KINEMATIC RECONSTRUCTION OF ttH, H $\rightarrow$ bb EVENTS AT THE LHC, AND SCIENCE OUTREACH THROUGH MULTIMEDIA BLOGGING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Shao Min Tan

May 2018

KINEMATIC RECONSTRUCTION OF ttH, H $\rightarrow$ bb EVENTS AT THE LHC, AND SCIENCE OUTREACH THROUGH MULTIMEDIA BLOGGING

Shao Min Tan, Ph.D.

Cornell University 2018

We embark on a parallel journey through physics, and the communication of physics. On one hand, we investigate a method of kinematic reconstruction suited for events containing top quarks produced in conjunction with other particles. After optimising the fitting process, we apply the reconstruction to $\mathrm{t\bar{t}H}, \mathrm{H} \rightarrow \mathrm{b\bar{b}}$ events in the single-leptonic channel. The reconstruction results in better estimates of particle momenta (including neutrino momenta) when applied to simulated events where the correct b quark permutations are known. However, when applied to fully-simulated MC datasets, it produces little change in the limits on the $\mathrm{t\bar{t}H}$ signal strength calculated using BDTs pre-trained on non-reconstructed events. In parallel, we study techniques for effective science communication for different audiences. In particular, we focus on my blog about the science of music and speech, which uses narrative elements, self-coded multimedia demos, and explanations at various levels of detail to present the subject in an appealing and understandable way. Finally, we briefly look at communication techniques for guided tours at CERN for the visiting public.

BIOGRAPHICAL SKETCH

Shao Min grew up in Singapore, and obtained a bachelor's degree from Carleton College in Northfield, Minnesota in 2012. Apart from doing physics, she writes a science blog and does outreach at CERN. She enjoys writing and playing music, learning languages, and dabbling in many other things. She also feels awkward writing about herself in third-person, and so begs leave to keep this biographical sketch short, and to end it, in fact, here.

To the inner child in all of us.

May we often let it roam;

May it nourish us in turn.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Jim Alexander, for his understanding, support and guidance over the course of my graduate studies. I am also grateful to my other committee members, Julia Thom-Levy and Maxim Perelstein, for their kindness and support. I'd also like to thank Aurelijus Rinkevicius for his knowledge and technical aid, as well as my Cornell groupmates Jorge Chaves, Jennifer Chu, Abhisek Datta, Susan Dittmer, Kevin McDermott, Nathan Mirman, Dan Quach, Louise Skinnari, Livia Soffi, André Sterenberg Frankenthal, Zhengcheng Tao, Jordan Tucker and Margaret Zientek, my friends Daniel Krefl, Dustin Anderson, Azar Eyvazov, Jonathan Gibbons, Jenny Goetz, Lipi Gupta, Lorien Hayden, Jack Jiang, Gabe Keller, Baldwin Mei, Jordan Moxon, Wee Hao Ng, Greg Rosenthal, Andrew Terwilliger and many others, my undergraduate professors Cindy Blaha, Arjendu Pattanayak, Bill Titus and Joel Weisberg, Cornell professors Paul Hyams, Colette Waldron, Kathy Selby, Damien Tissot, Bruce Lewenstein and Carl Franck, and of course, my parents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

THE STANDARD MODEL: THE THEORY OF NOT-QUITE EVERYTHING

Oh, you work at CERN! That's so cool! I studied physics in college, so I've always been awed by this kind of super high-tech research, but I never thought I'd actually meet a scientist from CERN...

Haha, we're just normal people, you know!

Normal? You must be super smart!

...

You know, I was afraid this was going to be a really long train ride, but I guess it's the perfect opportunity for me to ply you with questions! Didn't you guys make a big discovery a couple years ago? The Higgs boson, it was in the news...

Yeah! It was just before I joined the experiment though, so I wasn't there for it.

They call it the God particle and all that, but I never really understood what that was about.

That's because it gives mass to other particles.

What?

So the idea is, there's this thing called the Higgs field, and it's like viscous molasses that permeate all of space. And when particles move through space, they're slowed down by the molasses. Some particles don't interact with the molasses at all -- these are massless particles. And others do -- the more strongly they interact, the more mass they have.

That sounds really weird. I'm not sure I like the idea that I'm just wading through a bunch of molasses all the time.

Well, in a sense you are -- you're wading through a bunch of air molecules all the time, and you don't notice it! Anyway, the molasses thing isn't the most accurate analogy. But to go into more detail, I'd have to tell you a bit more about particle physics.

Go ahead!

What do you already know? You've heard of protons, neutrons and electrons, of course...

Yeah, the protons and neutrons are in the nucleus, and the electrons orbit around them, and that makes up the atom.

Yep, this model of the atom, with the positive charge at the centre and the negatively-charged electrons orbiting around it, was proposed in 1914. And they discovered the neutron about twenty years later. But this certainty about what matter was made of didn't last long. The field of quantum mechanics, with all its unintuitive

predictions, was being developed at around the same time, and that really shook things up...

1.1 The Electromagnetic Interaction

By 1923, the experimental evidence was incontrovertible [1]. Light exhibited very strange properties – it behaved like a wave at macroscopic scales, and like a particle at microscopic ones. Physicists reluctantly swallowed this fact, and set about to develop a theory that could reconcile the two halves of light's dual nature.

In 1927, Paul Dirac attempted to derive a linear-time equation of motion for a relativistic particle, by applying the Schrödinger Equation to the relativistic energy equation. The resulting Dirac Equation describes spin-$\frac{1}{2}$ particles such as the electron, and its solutions are represented by 4-component spinors ψ. Since these particles have half-integer spin, they are fermions which obey the Pauli Exclusion Principle.

Meanwhile, the idea of gauge invariance was gaining popularity among physicists. To make the Dirac Lagrangian invariant under a local phase transformation, one has to introduce an additional massless spin-1 vector field. This new field happens to be precisely the electromagnetic potential as described by the classical Maxwell Equations. Because the phase transformation involves a scalar phase, the electromagnetic interaction exhibits $U(1)$ gauge symmetry.

Under this theory, then, the electromagnetic field is quantised – it comes in little packets called photons. Photons are spin-1 particles, which means they are bosons which do not obey the Pauli Exclusion Principle. Two charged particles interact by exchanging photons, which play the role of mediators for the electromagnetic force.

1.2 Antiparticles

The Dirac Equation made a troubling prediction – solutions with negative energy! To account for this, Dirac made the rather fantastic suggestion that these negative-energy states are all filled with a perfectly uniform sea of electrons that we don't notice, so that the electrons that we observe must occupy positive-energy states. Whenever an electron is knocked out of the sea, however, the resulting "hole" would look like a positively-charged particle, of the same mass as an electron. Fortunately for Dirac, just such a particle was discovered in 1931, only a few years after he derived his equation [1]. The new particle was seen in the tracks left by cosmic ray particles in a cloud chamber, and was named the positron.

Richard Feynman and Ernst Stueckelberg later re-interpreted these negative-energy solutions to be positive-energy states of a different kind of particle, of the same mass but opposite charge, called an antiparticle. This eliminated the need for Dirac's invisible sea of electrons. Today, we know that all fundamental particles have an anti-version, with the opposite quantum numbers, though some particles (such as the photon) are their own antiparticle.

Ok, so they managed to derive the classical equations of electromagnetism, just by combining quantum mechanics and relativity. And at the same time they explained what a photon is. That's pretty cool.

Yeah, it is!

But there's still one thing I've always wondered about – you have all these protons together in the nucleus, and they're all positively-charged, so why don't they just repel one another and fly apart?

That was going to be the next part of my story! The simple explanation is that there must be another force, stronger than the electric repulsion between protons, holding the nucleus together. People named it the strong force.

Oh, I've heard of that. But if it's so strong, why don't we observe it in everyday life?

That's because it has a limited range -- its effects don't really extend outside the nucleus. It acts on the quar... -- have you heard of quarks?

Yeah, they're fundamental particles – protons and neutrons are made of them, I think?

That's right -- each is made up of three quarks. But we never see individual quarks in nature.

Then how do we know they exist?

By looking for patterns among the particles that we *do* see!

1.3 Quarks and the Strong Interaction

In the late 1940s, particle physicists began to detect a bunch of previously-unseen particles in cosmic rays. These particles were rather odd – they were produced plentifully, but seemed to decay rather slowly. With the advent of the particle accelerator in 1952, more and more of these strange particles were produced, of different masses and charges [1].

Sifting through the mess, physicists observed a pattern in how each particle was produced and how it decayed. They assigned to each particle a number called strangeness, which was conserved during their production, but not conserved during decay. In 1961, Murray Gell-Mann created a "Periodic Table" for these particles called the Eightfold Way, arranging them in geometric arrays according to their charge and strangeness (Figure 1.1). Just as the periodic table of elements illustrates that atoms have a substructure, the Eightfold Way could be explained by the fact that protons, neutrons and the new particles were made up of yet smaller constituents.

Gell-Mann named these smaller particles *quarks*. (Knowing from the first how the name would sound, he got the spelling from a line in James Joyce's *Finnegans Wake* where a drunken seagull mispronounces "quarts" while ordering three drinks.) Originally, three quarks were proposed, with rather less whimsical names – up, down, and strange. The particles detected so far consisted of either three quarks (baryons) or a quark and an antiquark (mesons). The proton and neutron are made of only up and down quarks (*uud* for the proton and *udd* for the neutron), while particles with the strange properties consist, naturally, of at least one strange quark.

Since then, three more quarks have been discovered, the last and heaviest of them

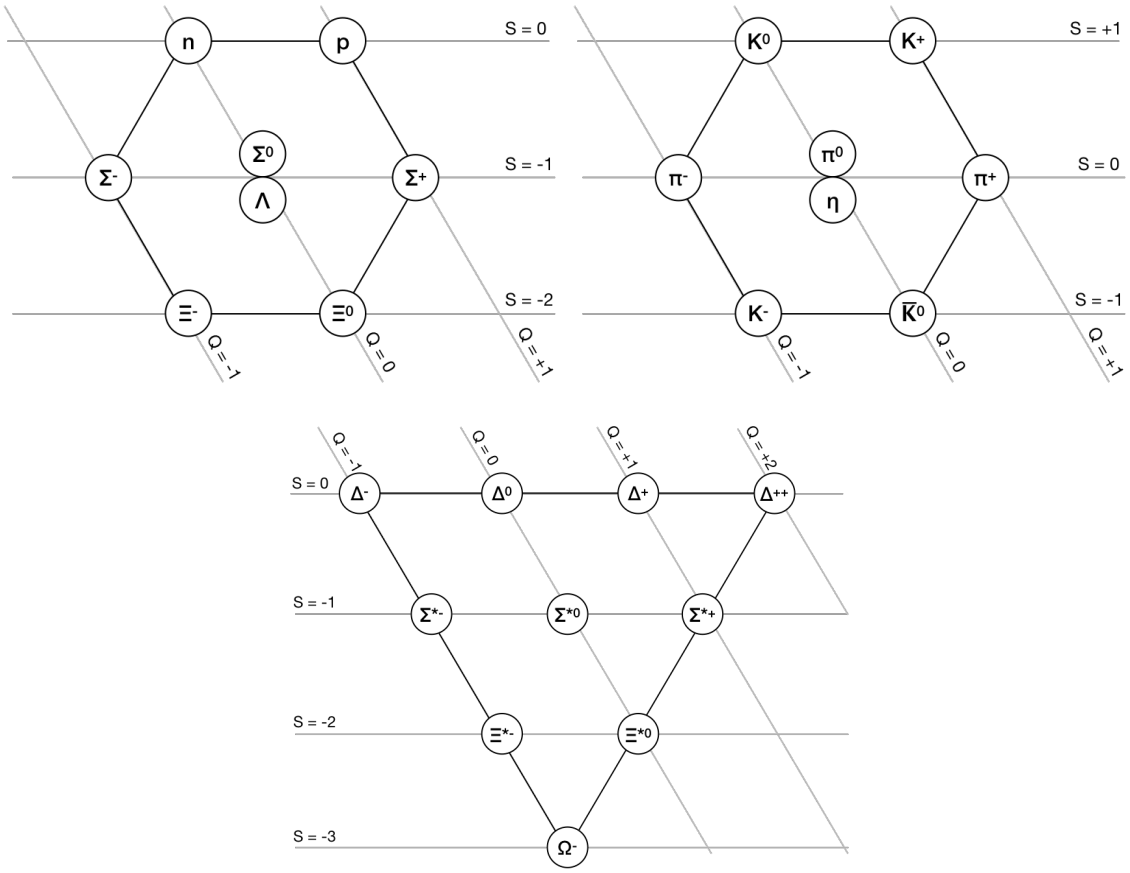


Figure 1.1: The baryon octet (top left), meson octet (top right), and baryon decuplet. The particles in each figure are arranged in a slanted grid based on charge Q and strangeness S.

(the top) in 1995. We arrange the quarks in three generations of two members each, as shown in Figure 1.2.

To explain why baryons such as *uuu* or *sss* do not violate the Pauli Exclusion Principle, Oscar W. Greenberg proposed that quarks have another property called *colour* – they can be red, green or blue. A baryon's three quarks each has a different colour, which means the quarks in, say, *uuu* are not identical, and the Exclusion Principle no longer applies. Baryons would also then fulfill the rule that all naturally-occurring particles are colourless (as with real colours, combining red, green, and blue gives a colourless object). Mesons, on the other hand, must have a quark of one colour and an anti-quark of the

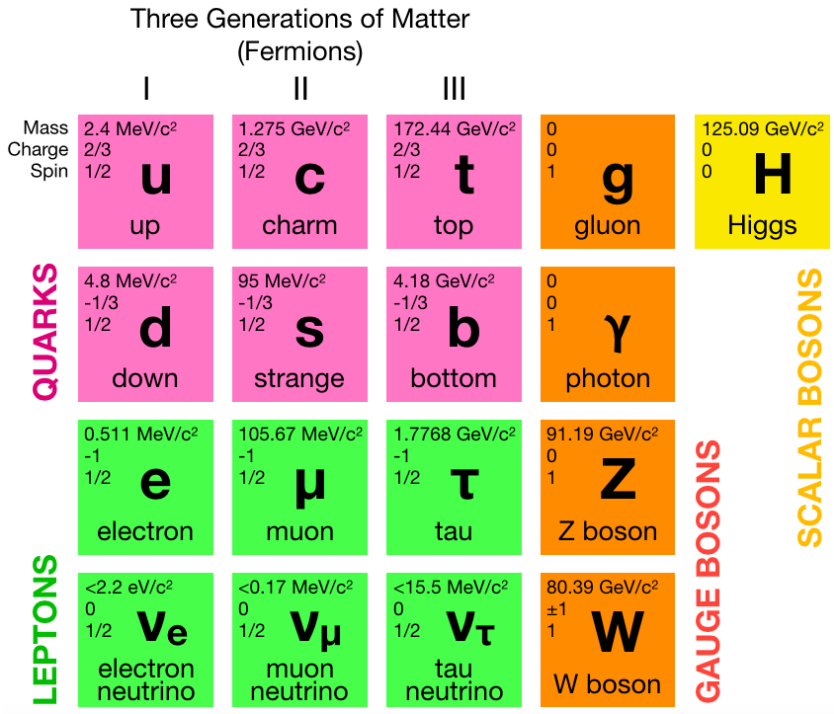


Figure 1.2: The elementary particles in the Standard Model.

corresponding anti-colour.

Quarks interact with one another via the strong force, which affects all objects which carry colour charge. It exhibits $SU(3)$ gauge symmetry on the three colours, and has 8 mediators called *gluons*. Gluons are massless, have a spin of 1, and themselves carry the colour charge (unlike the photon, which is not electrically charged).

We do not observe individual quarks in nature because of a phenomenon called *confinement* – quarks must always be grouped in baryons and mesons. When one tries to pull out an individual quark, it becomes energetically favourable to pop a quark-antiquark pair out of the vacuum, and the new quarks then combine with the separated quarks to form new baryons or mesons.

1.4 Leptons and the Weak Interaction

While observing the beta decay of radioactive nuclei in 1930, people noticed a weird problem. The electrons produced in the decay had a range of energies, even though the principle of conservation of energy dictated that they should have a fixed energy. Wolfgang Pauli suggested that another particle was produced in the decay, neutral and undetectable, that carried away the missing energy. It was named the *neutrino*, or "little neutral one".

The electron and the neutrino are known as leptons – fundamental particles that do not carry colour charge. We now know that there are three generations of leptons – the electron, muon and tau, each with its respective neutrino (as shown in Figure 1.2).

The force responsible for beta decay is known as the weak interaction. There are two kinds – neutral interactions, mediated by the Z^0 boson, and charged interactions, mediated by the W^+ and W^- bosons. Unlike the strong and electromagnetic interactions, the charged weak interaction can change the flavour of quarks – it couples up-type quarks to down-type quarks. (The strange particles introduced in section 1.3 were produced via the strong force, but decayed weakly; thus strangeness was conserved during production but not during decay.)

The weak force is also interesting in that it treats left- and right-handed particles very differently. In particular, the W bosons couple only to left-handed fermions and right-handed antifermions. In addition, the weak interaction violates parity and charge-parity symmetry.

In 1968 Sheldon Glashow, Abdus Salam and Steven Weinberg unified the weak

and electromagnetic forces into a single theoretical framework, that of the *electroweak* interaction. It is described by an $SU(2) \times U(1)$ gauge symmetry.

Unlike gluons and photons, the spin-1 weak mediators are massive. They were discovered in 1983 at CERN.

Hang on – the weak mediators are massive, you said. But earlier, when you were explaining the gauge symmetry thing for the electromagnetic force, you said that the gauge field had to be massless, otherwise the Lagrangian wouldn't be invariant...

Ha, you really *are* paying attention! That's the exact problem that led people to postulate the Higgs boson...

1.5 The Higgs Mechanism

We expect the $SU(2) \times U(1)$ gauge symmetry of the electroweak interaction to have four massless gauge fields: three W_μ^i's and a B_μ. Suppose we now introduce a new field, a complex scalar doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} ; \tag{1.1}$$

it has the Lagrangian

$$\mathcal{L} = (D^\mu \phi)^\dagger (D_\mu \phi) + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 , \tag{1.2}$$

where λ and μ^2 are positive, and D_μ is the covariant derivative which is invariant under the gauge symmetry:

$$D_\mu = \partial_\mu + igT^i W_\mu^i + i\frac{g'}{2}B_\mu \,. \tag{1.3}$$

Here, T^i are the generators of the $SU(2)$ group, and g and g' are coupling constants.

When considering a system's Lagrangian, which is related to its energy, a natural question is to ask where the minimum of the Lagrangian falls. After all, physical systems tend towards their minimum states, and we often do calculations by doing expansions about these minima.

For the Lagrangian in Equation 1.2, though, the minima do not fall in the centre of the coordinate system. We could shift the coordinate system so that its centre corresponds to a minimum, to make the expansion process more natural – but now the Lagrangian would no longer be symmetric. It has attained a vacuum expectation value (VEV) of $v = \mu/\sqrt{\lambda}$, and the Higgs doublet is now

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} , \tag{1.4}$$

where $h(x)$ is a new field.

This process of spontaneous symmetry breaking mixes the four massless gauge fields together, producing three massive fields (the Z_μ, W_μ^+, and W_μ^-) and one massless field (A_μ, the photon).

This mechanism for "giving mass" to gauge bosons, called the *Higgs Mechanism*, was proposed in the 1960s by multiple physicists, including Peter Higgs, Robert Brout and

François Englert. It can be expanded to fermions as well – when the quarks and leptons interact with the Higgs field, the term involving the constant v looks like a mass term – thus we also say that fermions obtain their masses via these interactions.

Of the original 4 degrees of freedom in the Higgs field, three are now "taken up" by the masses of the three massive bosons. The remaining one corresponds to a new boson, the massive spin-0 Higgs boson.

Aha, that's the one that was just found in 2012!

```
Indeed!  50 years after it was proposed...  Higgs and Englert finally
   got their Nobel Prize in 2013.
```

CHAPTER 2

WHY $t\bar{t}H$ PROCESSES?

So you found the Higgs boson. But you haven't shut down the LHC, of course. What are you still looking for?

Well, finding the Higgs boson was good news for the Standard Model. But we know that the model isn't complete -- there are still a bunch of problems with it.

It doesn't include gravity – that's one of the issues, right?

Indeed. Another one is something called the hierarchy problem.

2.1 The Hierarchy Problem

The mass of the Higgs boson m_H is related to the VEV by [2]

$$m_H = v\sqrt{2\lambda}\,. \tag{2.1}$$

The Standard Model also tells us that the physical mass of the Higgs boson is a combination of the bare mass m_0 and loop corrections: $m_H^2 = m_0^2 + (\Delta m_H^2)$ [3]. A first-order loop correction from a fermion that couples to the Higgs is shown in Figure 2.1a, and gives us a quadratically divergent correction term. We could apply a cutoff to the loop integral, but we would expect this cutoff to be at the Planck scale, $m_P \approx 10^{19}\,\text{GeV}$. This results in a correction term Δm_H^2 of order $(10^{19}\,\text{GeV})^2$.

But we know that m_H is of the same order of magnitude as v from Equation 2.1 (since we want λ to be of order 1 in order for our theory to remain perturbative). v is known experimentally to be about 246 GeV, so m_H is of order 100 GeV. This means that the bare mass parameter m_0^2 would have to be of the same order as Δm_H^2, i.e. $(10^{19}\,\text{GeV})^2$, and be fine-tuned in such a way as to miraculously cancel out Δm_H^2 to produce a physical mass of order 100 GeV. A theory that requires such a precise coincidence just doesn't seem very *satisfying*.

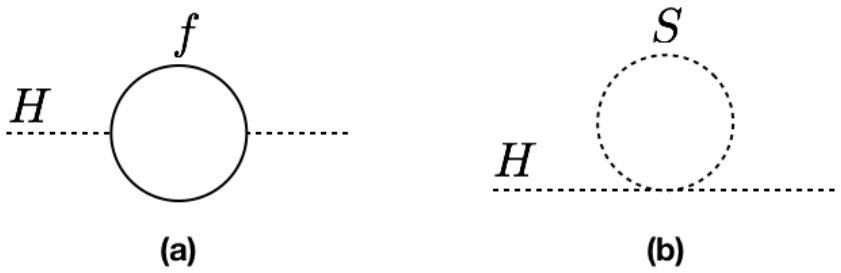


Figure 2.1: (a) Fermion loop correction to m_H. (b) Scalar loop correction to m_H. Figure adapted from [4].

2.2 Beyond the Standard Model

Various models have been proposed to solve the hierarchy problem, of which the most well-known is Supersymmetry (SUSY). This model relates fermions and bosons, so that each SM fermion has a bosonic superpartner, and each boson has a fermionic superpartner. Each particle's superpartner is able to generate loop corrections to m_H that cancel out the original particle's contribution. In the fermion example studied above, for instance, the fermion would have a scalar superpartner $\tilde{f}$ which contributes the diagram in Figure 2.1b. This contributes a correction term that has a relative minus sign compared to the fermion's correction term [3], and fine-tuning is no longer needed.

Crucially, such Beyond-the-Standard-Model (BSM) theories predict different values for the coupling of the Higgs to various particles, compared to the Standard Model. These couplings can be measured at particle accelerators.

I see – so you measure these couplings, and based on their value you can choose the theoretical model that comes closest in its predictions.

Well, it's a bit more complicated than that, because these values are really hard to measure -- but that's the general idea.

How exactly do you measure these couplings? It seems very abstract...

We don't measure them directly. Instead, we look at the *rate* at which certain processes happen. The rate is dependent on the coupling strength, you see.

And is this what *you* do, personally?

Yep. I look at one particular way in which the Higgs is produced -- with two top quarks.

2.3 The $t\bar{t}H$ Production Channel

At the LHC the dominant Higgs production channel is gluon fusion (Figure 2.2, left), but the channel involving the production of a Higgs boson in conjunction with a top

quark pair (Figure 2.2, right) has some advantages over gluon fusion despite its much smaller cross-section [5].

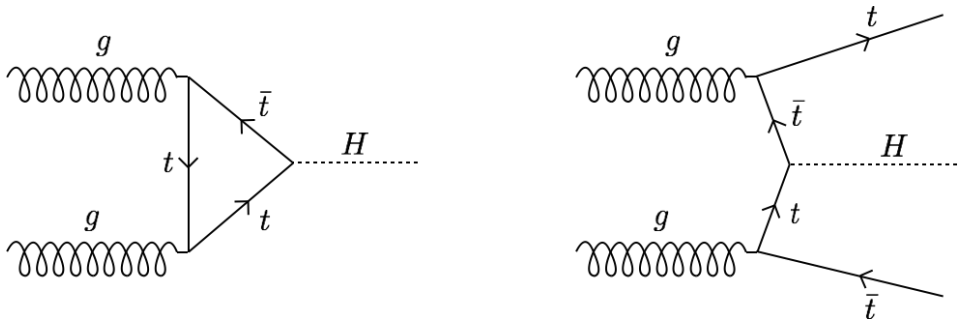


Figure 2.2: Feynman diagrams showing gluon fusion production of the Higgs (left) and Higgs production in conjunction with a $t\bar{t}$ pair (right).

Firstly, the $t\bar{t}H$ production mode allows us to probe the top-Higgs Yukawa coupling directly [5]. (Because the Higgs, at 125 GeV, is lighter than the top quark, we cannot measure this coupling by searching for Higgs decays to top quarks [6].)

In addition, the $t\bar{t}H$ channel allows us to access the bottom-Higgs coupling, since the dominant decay mode of the Higgs at 125 GeV is to a $b\bar{b}$ pair. Searching for the $H \rightarrow b\bar{b}$ decay of a Higgs produced from gluon fusion is unfeasible due to large backgrounds, but $t\bar{t}H$ or Higgs production in conjuction with a vector boson have significantly smaller backgrounds for this decay channel [7]. $t\bar{t}H$ is thus a very useful production mode to study, giving a handle on the Higgs coupling to both the top and the bottom quarks.

CHAPTER 3

A BRIEF HISTORY OF SCIENCE COMMUNICATION

Geez, I have a crick in my neck from turning left to face you this whole time... What do you say we change seats?

Good idea, my neck hurts too...

*****

Hey, I'm getting dry from all that talking, so I'm going to declare that the left-hand seat is the hot seat. Whoever sits in that has to do the storytelling... So it's your turn now! You said that you're a grad student too?

Yeah, I study science communication.

This is embarrassing, but I didn't realise that that was something that people actually studied...

You think of science outreach as something that scientists do in their free time? That's a pretty common view among scientists, I think.

I think many scientists see it as something that they'd rather not do if they had the choice! But in recent years it seems that there's been an increasing focus on outreach among funding agencies -- it's become almost a requirement, if you want to get grant money.

Believe it or not, this situation has actually been around for quite a while! Think about it – before the 19th century, science wasn't even really a thing. People studied the world, of course, and made discoveries in astronomy and biology and physics. But most of these natural philosophers, as they were then called, weren't professional.

```
Ah, right, they were kind of like artists -- they had to have patrons
    among the rich and noble.
```

Yup. And they did communicate what they were doing; at the very least, to keep their sponsors happy. Galileo, for instance, published books and essays in which he explained his astronomical theories, in a way that a layperson would understand. His works were often dedicated to his patrons, and some of them caused a bit of an uproar in the church.

```
When did science become a professional activity, then?
```

Around the 19th century – the term "scientist" was coined in 1833 (8). But the process had already been set in motion earlier...

3.1 The Age of Natural Philosophy

The first scientific academy was the Accademia dei Lincei, founded in 1603. It was followed by the Royal Society in 1660, and the Academie des Sciences in 1666 [9]. These societies played an important role in increasing communication and collaboration between scientists. Scientific writings at the time were usually in Latin, which allowed them to be shared across national borders, but restricted access to the more educated segment

of the population [10]. With the rise of national societies, however, scientists began to write in their own language [9].

While the sharing of ideas between *scientists* might not seem like our idea of science communication, scientific writings of the day were not the dense jargon-filled treatises that we see in journals today. Science was not yet highly specialised, and publications were relatively understandable for the well-educated layperson. The non-professional nature of science meant that scientific discourse was often informal, taking place in cafés and other public places, and open to anyone with an interest in the subject [10].

The need to make science appealing to the general public existed even back then. Before the 19th century, most scientific research was not done at universities. (These educational institutions tended to focus on classical subjects such as theology, philosophy, and Latin; the only scientific subjects to be taught were medicine and pharmaceutical science.) To fund their endeavours, many scientists turned to giving public lectures and demonstrations. Since they depended on their audience for a living, it was essential to keep them interested. Lecturers would focus on the applied sciences, rather than abstract subjects such as mathematics; and they would emphasise the benefits of new technology [10]. Public demonstrations were often equal-parts entertainment and education.

An early example of a scientific celebrity was British chemist and inventor Sir Humphry Davy. Davy was engaged by the Royal Institution, which had been set up in 1799 with the purpose of diffusing knowledge about science and technology. He gave public lectures, directed the chemistry laboratory, and edited the institution's journals in exchange for a salary [11]. Davy was a charismatic speaker, and his lectures often included crowd-pleasing elements such as spectacular (and sometimes dangerous) demonstrations of chemical reactions. He would also link his work to poetry and religion, in an attempt to

better appeal to his audience (especially the women – or so he saw it). Davy's lectures were so popular that the road outside the Royal Institution had to be made into a one-way street during his demonstrations to relieve congestion [10]!

Figure 3.1: Cartoon of a Royal Institution lecture on pneumatics, 1802. Humphry Davy is holding the bellows. Figure from [12].

3.2 Museums and Exhibitions

The 19th century also saw the rise of a different form of science communication – that of museums and exhibitions. Following in the wake of the industrial revolution, technological expositions began to crop up around Europe and the United States. The first World's Fair was named *The Great Exhibition of the Works of All Nations*, and took place in London in 1851. It included exhibits that showed how machines worked, such as how cotton was spun and made into cloth [13]. Scientific instruments and new inventions

were also displayed, alongside non-scientific exhibits such as jewellery and art. This focus on industry and technology would define the World's Fairs until the 1930s.

The organisers of the Great Exhibition set up special railways to bring people in to the venue, and printed different ticket types with varying prices depending on the date of visit [9]. This allowed people of all social milieus to attend. The trend of making information accessible to the lower classes started to pick up, as private galleries gave way to public collections. This change took place slowly, however – when the British Museum first opened in 1753, it admitted only 15 people a day, all of whom were first subject to a background check that could take months [10]! As visitor numbers to other museums increased, commentators were surprised that the crowds were relatively well-behaved, and did not devolve into angry, drunken mobs. Museums would even specially encourage women to attend exhibitions, believing that their presence would inspire the men to behave in a more civilised manner [10].

The shift from private to public galleries meant that museum collections had to be presented differently. Before, a guide would be employed to show around a few visitors while providing commentary. Now, curators had to provide written labels, organise the displays, and think about how to arrange the exhibits to guide the flow of visitors through the rooms.

3.3 Professional Science

In the later part of the 19th century, science became increasingly professional. Universities began to involve themselves in scientific research, and fields of study became more specialised. This meant that new research could no longer be easily understood by

non-experts.

A distinction began to emerge between publications aimed at scientists and those aimed at laypeople. The former became terse and more difficult to read – hardly lending themselves to a relaxed perusal during one's leisure hours [10].

The roles of scientist and science communicator were also diverging. Scientists would concentrate on research, while the job of communication fell to a different set of people. These science communicators played the role of a bridge between professional scientists and laypeople, using more accessible forms of writing such as science fiction [10].

As "popular science" became a genre, scientists began to form their own identity, seeing themselves as an exclusive group of expert practitioners. As Peter Broks [8] put it, 'excluding the public became a defining feature of what it meant to be "scientific"'.

3.4 Science Journalism

Ah, that's interesting. It seems that the audience that people were communicating to was constantly changing. And even the distinction between scientists and the public was a relatively recent thing.

Yes, but the changes didn't stop there. Throughout the 20th century, the relationship between scientists, science communicators and the public continued to evolve, as can be seen by how science was portrayed by the media over the decades.

In the 1940s, following the second World War, the west experienced a period of

rapid growth and industrialisation. The atmosphere in science journalism reflected this general optimism – Boyce Rensberger [14] termed this the "Gee Whizz Age" of science writing. Feeding off the curiosity and enthusiasm of the public, journalists wrote about the marvels of science and technology, speaking of scientists with a respectful tone.

All this changed in 1962, when marine biologist Rachel Carson published her book *Silent Spring*, which described the damaging effects of the pesticide DDT on wildlife. This was significant because DDT had previously been seen as a technological wonder. The tone of science reporting began to change, with journalists starting to question the ethics and techniques of science. Not all science reporters were comfortable with this "watchdog" role – some felt that it was their duty to work with scientists in order to produce good writing, and that it was important to retain the trust of the scientific community [15].

The idea that science reporters should act as "watchdogs" for society was strengthened in the 1970s, after a series of high-profile incidents associated with new technology [15]. An example was the 1979 Three Mile Island accident in Pennsylvania, the worst accident in the history of commercial nuclear power in the United States. The social and environmental impacts of science and technology were suddenly brought to the fore of public consciousness, and science reporting took on a more critical tone.

Over the next two decades, science journalism experienced a boom. Many major newspapers ran science sections, and new specialist science magazines were launched. Numerous scientist celebrities also rose to fame during this period, appearing on television and authoring books. A notable example is Stephen Hawking, whose 1988 book *A Brief History of Time* would sell 10 million copies in 20 years [15].

3.5 Public Understanding of Science

In 1985, the Royal Society published a report named *Public Understanding of Science*, later known as the Bodmer Report. It described the current state of science communication in bleak terms. Scientists had become too detached from the societal implications of their work, and had retreated too deep into the cocoon of their laboratories. The few that did participate in outreach activities were seen as inferior scientists by their colleagues. This situation, the report said, could endanger scientific funding [16].

The suggestions put forth by the Bodmer Report marked a turn in scientists' attitudes towards outreach in the UK. Communication was now encouraged; committees and organisations were set up and funded to disseminate scientific knowledge. Programmes included science book prizes and funding for science communication practitioners. In particular, the movement encouraged young and early-career researchers to be trained in and to participate in outreach, in contrast with the earlier view that only senior scientists were qualified to be spokespeople for their field [16].

3.6 The Deficit Model

That sounds pretty fantastic.

It does sound good – but unfortunately, it didn't work!

It didn't work?

Yeah, over the years they did a bunch of surveys measuring how much the public knew about various scientific topics. They found that the results from the 1988 and 1996 surveys showed little difference in scientific knowledge, despite 8 years of Public Understanding of Science efforts in between (16).

What went wrong, then?

The problem was something called the Deficit Model. It was the assumption, in all the science communication efforts so far, that science communication was all about filling the public with science information. That outreach involved the flow of knowledge from the knowledgeable scientist to the deficient layperson, who was just sitting about like an empty vessel, ready to take in new information.

Well... if you put it like that it does sound rather elitist, but... isn't it true that scientists *do* have superior knowledge?

In many cases, perhaps – but not always. Here's an example – in 1992, a study was published about the interaction between nuclear scientists and sheep farmers in Cumbria. The scientists had been studying the effect of the Chernobyl nuclear accident on the sheep, but they weren't interested in listening to the farmers' specialist knowledge about their animals. As a result of their arrogance, their credibility in the eyes of the farmers was damaged, and some of their experiments also eventually failed (17).

Oh, that's silly.

Another problem with the Deficit Model is that it assumes that once people know more about science, they'll be all onboard and enthusiastic and supportive of

science. But that isn't necessarily the case. People who know more about a subject can sometimes be more critical of it. For instance, college-educated Republicans are actually *more* likely to be skeptical of global warming than less educated ones (18).

```
Really!  That's surprising.
```

So what people found was that it's very important to take social context into account when doing science communication, especially if the subject is something that affects people directly. Scientists might possess the scientific facts – but the public may have their own local knowledge of something, or a personal interest in it, that you ignore at your own peril.

3.7 Public Engagement and Citizen Science

Since the 2000s, science communication has moved in a new direction – that of *engaging* the public. Instead of increasing public *understanding* of science through a one-way flow of knowledge, we now encourage a two-way exchange of ideas between scientists and the public. This idea makes sense in light of a 1999 UK study, which showed that, while the public thought that science was fascinating and made their lives better, they also mistrusted scientists – they thought that scientists did not pay enough attention to the risks of research, and that regulations would not keep them from doing what they wanted behind closed doors. This indicated that, instead of emphasising the wonders of science, communicators should focus on gaining public trust by giving them a say in how science is done [19].

An example of how such a dialogue could take place is the consensus conference. Here, lay citizens are given an extensive briefing of a certain topic, and asked to evaluate new scientific methods and issues [16]. However, in practice, a direct pipeline from public to science policy is often not so easily established. More common are other activities that do not give the public direct influence over policy, but still attempt to engage them. For example, the Café Scientifique has people gather in a bar to meet a local scientist, to discuss the implications of research [19]. Lower on the engagement scale is the science shop, where the public can go to find information about issues affecting the local community [16].

But wait, I haven't experienced any of these citizen panel things. Most of the outreach events I go to still seem to follow the filling-empty-vessels-with-facts model -- albeit, it's done in an interesting way.

Ah, but you're in particle physics. It depends a lot on the field – different fields lend themselves to different methods. Controversial subjects that have a direct impact on people's lives, like genetic modification or environmental studies, might be more suited to citizen involvement. But for something like particle physics, which is pretty far-removed from daily life, and which has more of a tendency to inspire wonder, a more one-way flow of information could be expected.

Yeah, I guess the same applies to those David Attenborough documentaries -- you're just supposed to sit there being filled with wonder at the beauty of nature. I suppose it also depends on what your goal is for a particular outreach activity. Even within a certain field, you can have outreach that aims to simply inform and inspire, while

other programmes involve the public more directly.

Certainly – when doing science communication, the most important thing is to keep your goal and your audience in mind. Nobody is saying that science centres should stop designing fun exhibits for children, and instead get their visitors to sit around discussing the implications of bioengineering. We certainly shouldn't consider one-way transmission of knowledge as *inferior* to two-way models, nor should we discount the importance of scientific knowledge. Your approach depends on what you're trying to do.

CHAPTER 4

AT THE LARGEST MACHINE IN THE WORLD

Time to switch seats again? OK...

So what's it like working at CERN? Must be cool to sit next to these big machines every day!

Most of us don't sit next to them -- the LHC is underground, and you can't hang out there while it's running, because of the radiation. You know, whenever someone visits me at CERN, they're always surprised how the campus looks. They expect all this really high-tech cutting-edge stuff, but it's mostly a bunch of blocky post-war style buildings with door handles that are falling off...

Haha, but the machines are high-tech, at least?

They fulfill every possible cliché of high-techness. The machine is this huge ring that's 27 kilometers in circumference, and lies about 100 metres beneath the French and Swiss countryside near Geneva. We use it to collide beams of protons or lead ions at energies never before achieved.

So there's two beams going in opposite directions?

Yep, and they collide at four points on the ring, corresponding to four detectors. The one I work on is called the Compact Muon Solenoid,

or CMS. The others are ATLAS, LHCb, and ALICE.

4.1 The Large Hadron Collider

4.1.1 Physical Design

The Large Hadron Collider (LHC) is designed to collide beams of protons with energies of up to 7 TeV, corresponding to a centre-of-mass collision energy of 14 TeV. The protons in each beam are grouped into bunches of about 1.15×10^{11} protons each, and each beam is designed to hold up to 2808 bunches. The bunch spacing, or time that elapses between successive collisions, can be as low as 25 ns.

Beams are accelerated by alternating electric fields produced by radio frequency cavities. To bend a beam, on the other hand, we use superconducting dipole magnets with fields of up to 8.3 T. There are also quadrupole and higher-order magnets whose role is to keep the beam focused. The magnets are cooled with superfluid helium to their operating temperature of 1.9 K – that's lower than the ambient temperature of the universe! Figure 4.1 shows the cross-section of a dipole magnet, showing the two beam pipes and the coil magnet structures surrounding them.

To get the beams up to speed, they are passed through a series of increasingly larger accelerators, whose infrastructure is partly built on CERN's older, retired machines. Figure 4.2 shows the accelerators in the CERN complex. Protons are first produced by stripping the electrons off the hydrogen atoms in a gas canister. They are then accelerated by the Linac2 linear accelerator to 50 MeV, then by the PS Booster to 1.4

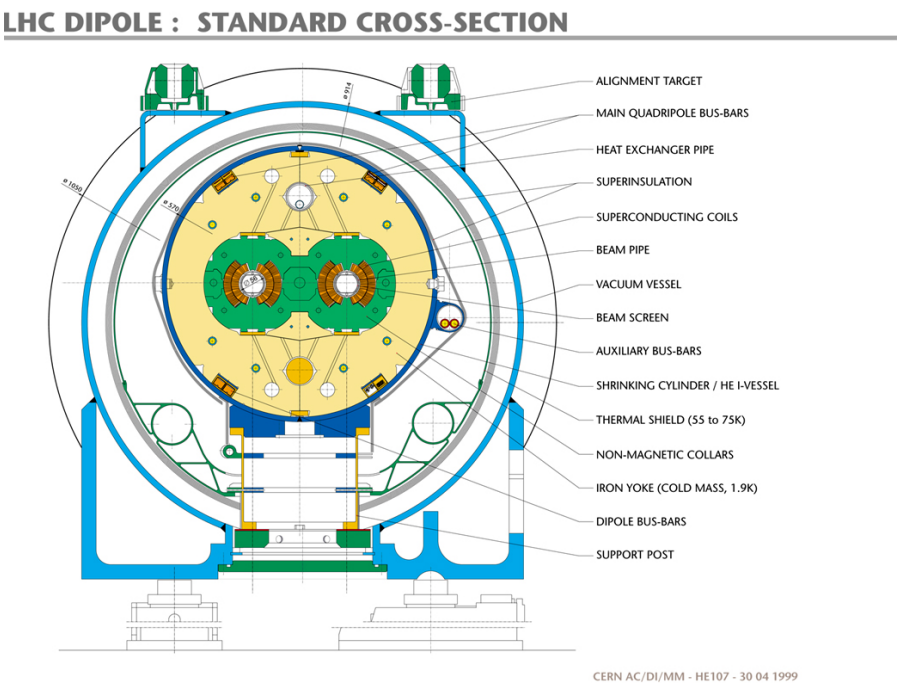


Figure 4.1: Cross-section of LHC dipole magnet. Figure from [20].

GeV, then by the Proton Synchrotron to 25 GeV, by the Super Proton Synchrotron to 450 GeV, and finally injected into the LHC to be accelerated to full energy. Beams can circulate for about 10 hours; after this too many of the protons will have been "used up" in collisions, and a fresh beam needs to be produced.

4.1.2 Operation

The instantaneous luminosity $\mathcal{L}$ of a collider allows us to calculate the rate at which we expect to see a process of a certain cross-section σ:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \mathcal{L}\sigma\,. \tag{4.1}$$

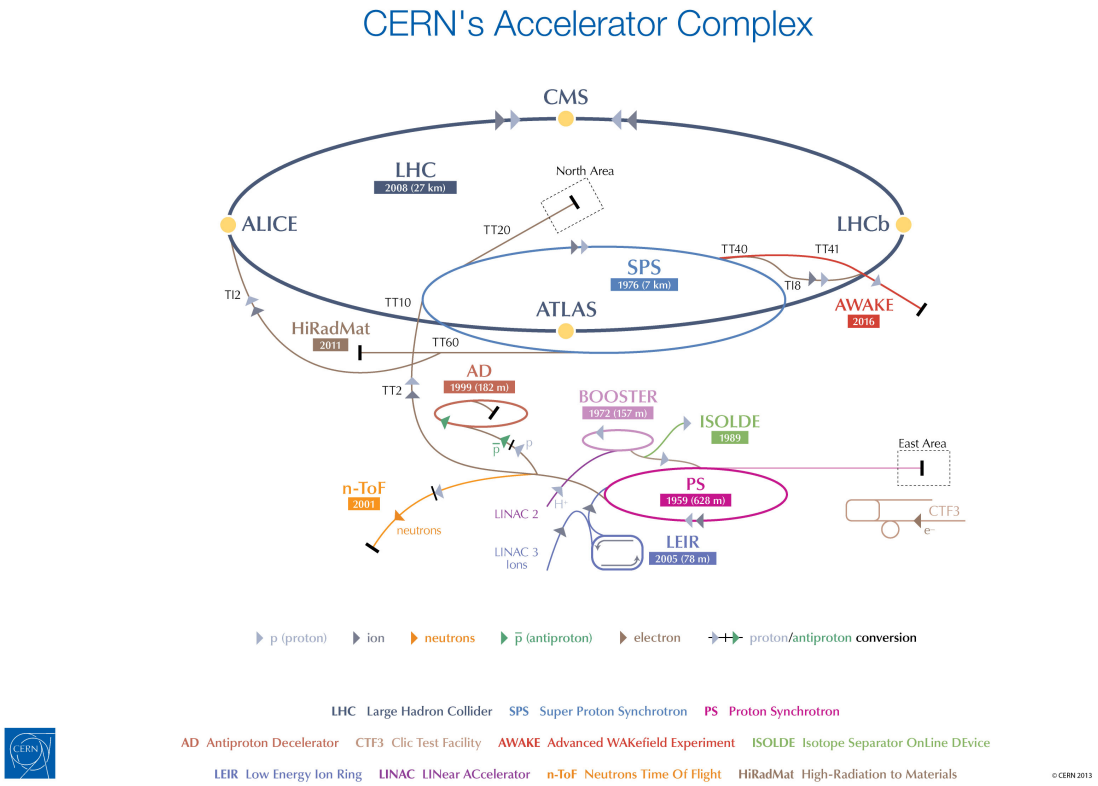


Figure 4.2: The CERN accelerator complex. Figure from [21].

Among other factors, $\mathcal{L}$ depends on the bunch size, number of bunches and circulation frequency of the beam, and varies from day to day and from run to run. The integrated luminosity $\mathcal{L}_{\text{int}}$ is a measure of how many collisions have occurred over a period of time:

$$\mathcal{L}_{\text{int}} = \int \mathcal{L} \mathrm{d}t \,. \tag{4.2}$$

This quantity is usually measured in inverse femtobarns (fb^{-1}).

The LHC started collisions in 2010, and ran at 7 - 8 TeV with a bunch spacing of 50 ns until 2012, a period of operation known as Run I. It was shut down in 2013 and 2014 to allow for upgrades and repairs, and re-started in 2015 for Run II. The centre-of-mass

energy was increased to 13 TeV, and the bunch spacing reduced to 25 ns. Figure 4.3 shows the cumulative integrated luminosity delivered to CMS as a function of the date, for each year of the LHC's operations.

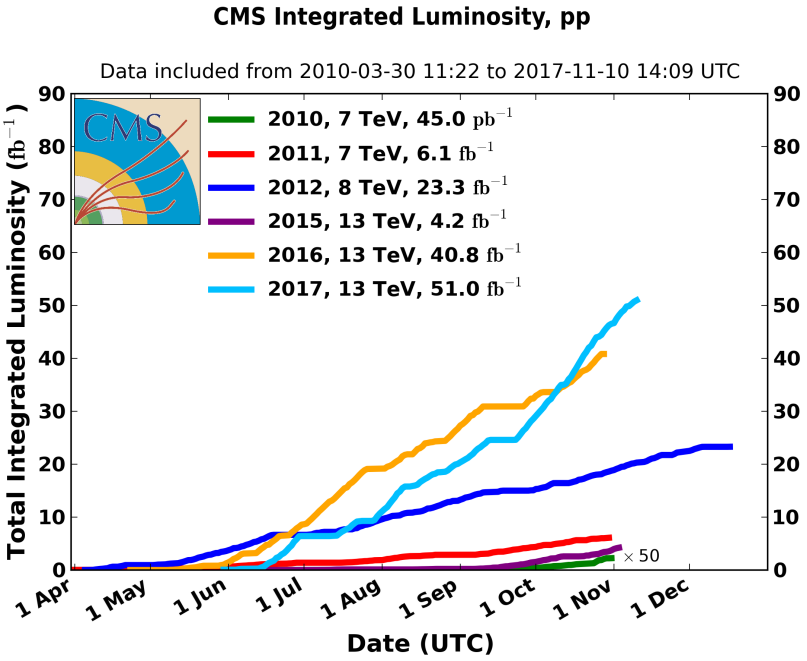


Figure 4.3: Integrated luminosity delivered to CMS as a function of date, for each year of LHC operation. Figure from [22].

As the instantaneous luminosity increases, so does the probability of getting a collision in a bunch crossing. At its design luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$, the machine averages approximately 20 collisions per bunch crossing. This phenomenon of multiple interactions is called pileup, and it complicates the process of reconstructing events.

4.2 The CMS Detector

So you said earlier that the beams collide at four points around the LHC ring, where the detectors are. How do you make sure they only collide at those points?

Remember the cross-section of the dipole magnet you saw in Figure 4.1? There you could see that there are two beam pipes, for the two beams going in opposite directions. These beam pipes cross at the four collision points, so it's only possible for the beams to collide there.

Ah. And what happens once you get a collision?

The detector measures the particles that fly out from the collision. CMS has many layers of detector components, each of which is responsible for measuring different quantities and different particles. It's a bit like a cylindrical onion.

The beam pipe of the LHC passes through the central axis of the CMS cylinder, with the interaction point positioned in the centre of the detector. A superconducting solenoid magnet with a field of 3.8 T sits between the inner detectors and the outer muon chambers. The magnetic field is aligned with the beam pipe, and causes particles flying out of the collision to bend in the cross-sectional plane. This allows us to measure their momenta.

The coordinate system of CMS has the x-axis pointing toward the centre of the LHC ring, the y-axis pointing upwards, and the z-axis pointing parallel to the beamline. In

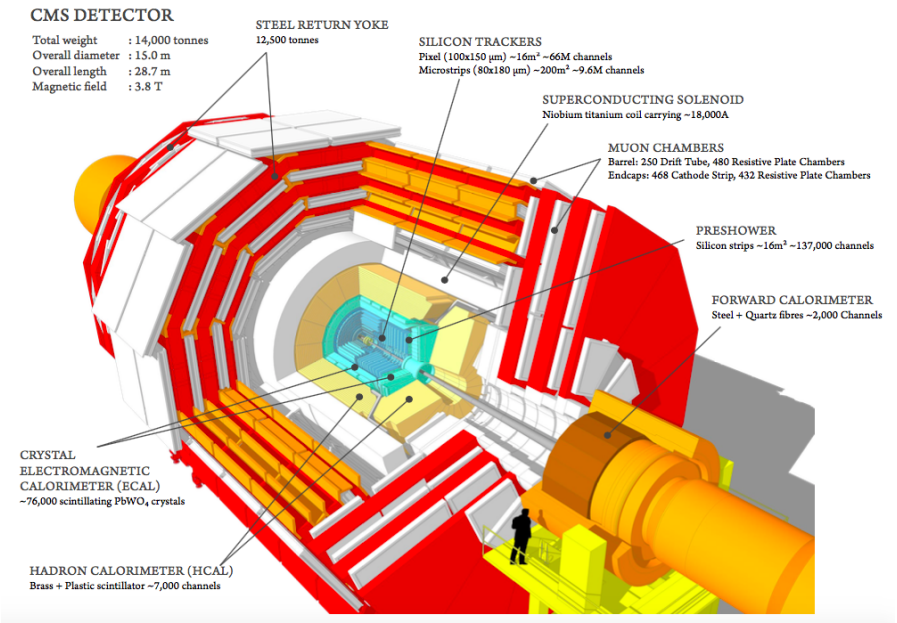


Figure 4.4: A cutaway view of the CMS detector, showing the various detector components. Figure from [23].

practice, we usually express particle momenta in cylindrical coordinates p_T, ϕ and η. p_T is the momentum in the transverse (x-y) plane, and ϕ is the angle from the x-axis in the x-y plane. Pseudorapidity, or η, is given by

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{4.3}$$

where θ is the polar angle from the z-axis. Pseudorapidity is preferred to θ because the difference in rapidity between two points is roughly invariant under Lorentz boosts in the z-direction. Finally, the last defining component of a particle's 4-momentum is its mass m or energy E.

4.2.1 The Inner Tracker

Directly enveloping the collision point is the inner tracker, whose role is to map out the paths of charged particles. It needs to do so while obstructing them as little as possible, so that their energies can be accurately measured in the next detector layer.

The tracker is made of two parts. The inner component extends out to 10.2 cm in the radial direction and 46.5 cm in the longitudinal direction. It consists of 66 million silicon pixels of size $100 \times 150 \mu m^2$, and has a resolution of 10 - 20 μm [24]. It is laid out in a cylindrical configuration, with three layers in the barrel (the curved part) and two in the endcaps (the two disc-shaped parts). The outer component extends out to 1.1 m, and is made up of 9.6 million silicon strips of thickness between 320 and 500 μm, and of dimensions ranging from 10cm $\times$ 80μm to 25cm $\times$ 180μm. It is arranged in 10 layers in the barrel and 12 discs in each endcap, and has a resolution between 23 and 53 μm. Between the barrel and endcaps, the tracker has an acceptance up to a pseudorapidity of $|\eta| < 2.5$ [25].

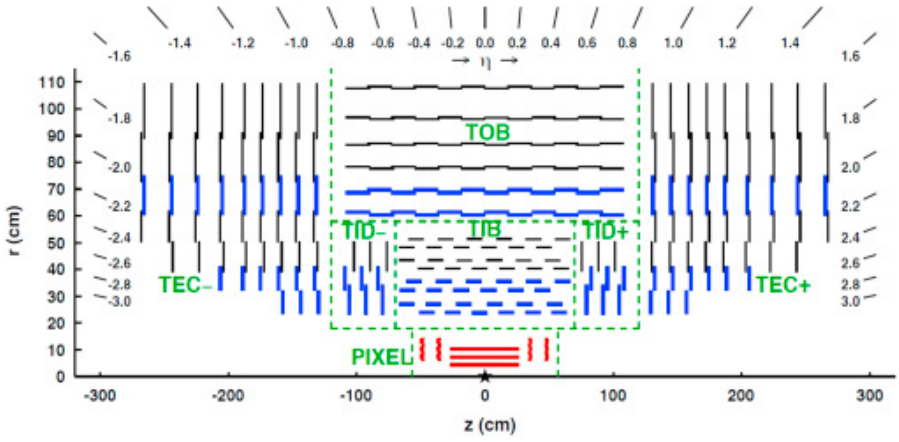


Figure 4.5: Layout of the CMS tracker in the r-z plane, showing both the pixel and the strip detectors. Figure from [26].

As particles travel through the tracker, the silicon pixels and strips produce electrical signals that are transferred out by 75 million electronic read-out channels [27]. Because it is the closest layer to the collision point, the tracker is exposed to high amounts of radiation. Though the tracker was designed to be radiation-resistant, the damage caused to it during LHC operations limits the lifetime of its components to between 2 and 10 years, depending on how close the layer is to the centre [25].

4.2.2 The Electromagnetic Calorimeter

The next layer of CMS is the electromagnetic calorimeter, or ECAL, whose job is to measure the energy of electrons and photons. The ECAL is made of lead tungstate ($PbWO_4$) crystals – 61200 in the barrel and 7324 in each of the endcaps. Passing electrons and photons cause the crystals to scintillate and give off light, which is then read out by avalanche photodiodes (APDs) and vacuum phototriodes (VPTs) in the barrel and endcaps respectively. The barrel covers the pseudorapidity range $|\eta| < 1.479$, and the endcaps $1.479 < |\eta| < 3.0$. In addition, there is a preshower detector in front of each endcap, made of silicon strip sensors and lead radiators. Its purpose is to improve the position measurement of electrons and photons, as well as to detect neutral pions.

Lead tungstate was chosen for its advantageous physical properties. Its short radiation length means that the ECAL can remain relatively compact while being able to absorb most of the energy from photons and electrons. It has a small Molière radius – a measure of the size of electromagnetic showers caused by a passing photon or electron – which gives it a good position resolution. In addition, its scintillation time is short – about 80% of the light is emitted in the bunch-crossing time of 25 ns. This means that there is less overlap in the signals produced by successive bunch crossings. Lead tungstate also has

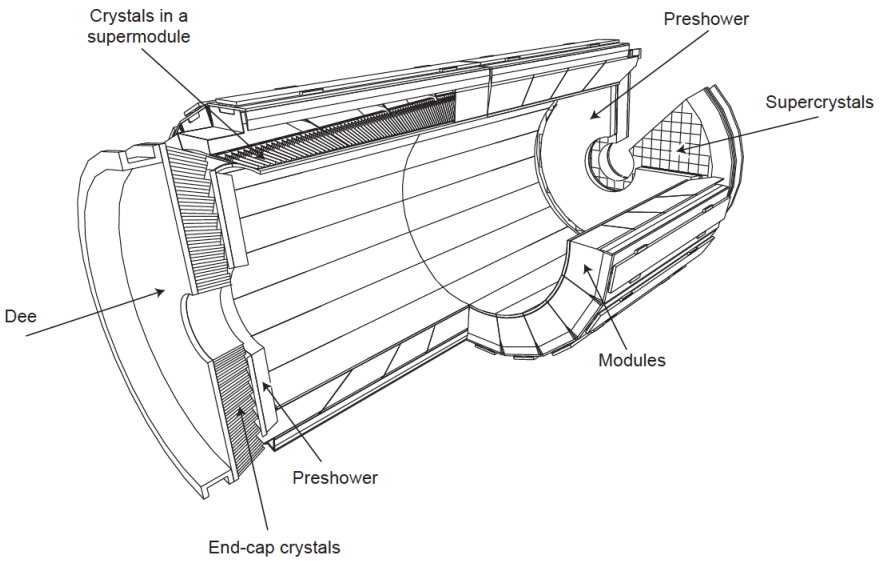


Figure 4.6: Schematic of the ECAL. Figure from [28].

the advantage of being resistant to radiation [25].

4.2.3 The Hadron Calorimeter

The hadron calorimeter, or HCAL, makes up the next layer of the CMS onion. Its purpose is to measure the energy of hadrons. The HCAL is made up of four parts: the barrel (HB), endcap (HE), outer calorimeter (HO) and forward calorimeter (HF), as shown in Figure 4.7.

HB and HE consist of alternating layers of brass plates and plastic scintillating tiles, and the scintillating light is read out by wavelength-shifting fibres embedded in grooves in the scintillator. HB covers a pseudorapidity range of $|\eta| < 1.3$, and HE covers $1.3 < |\eta| < 3$. HO lies outside the solenoid magnet in the pseudorapidity range $|\eta| < 1.3$, and captures any energy not absorbed by HB. It consists of a thick iron layer and one or

two scintillator layers, depending on the η position [25].

Finally, the HF covers the high-rapidity region, where particle flux is the highest. It is made of quartz fibres (which have high radiation hardness) embedded in grooved steel plates, and generates a signal consisting of Cherenkov light.

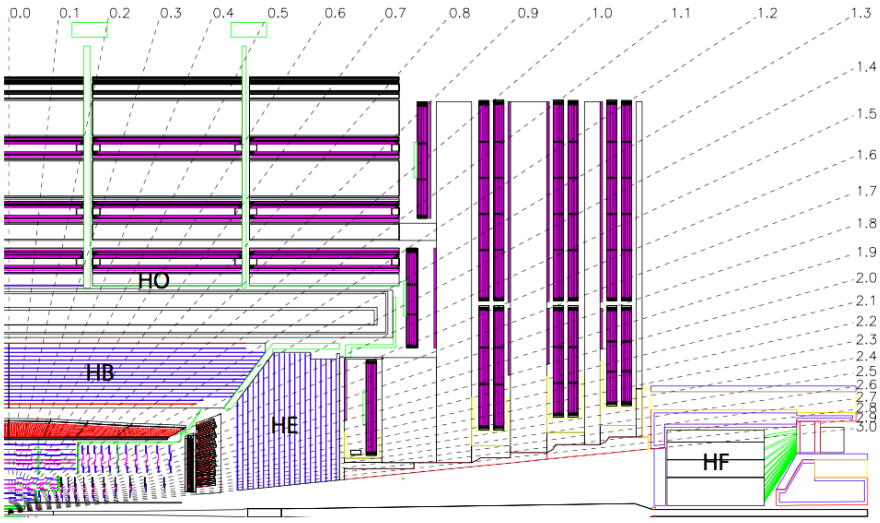


Figure 4.7: Schematic of the HCAL, showing the four detector components. Figure from [25].

4.2.4 The Muon System

The last detector, extending out beyond the solenoid magnet, specialises in measuring muons, which do not leave much energy in the inner calorimeters. An iron return yoke wraps around the magnet and is interleaved with the muon chambers, providing a magnetic field. CMS has three types of gaseous particle detectors in the muon system. When muons pass through, they knock electrons off the gas atoms, and the electrons are detected by wires or other conducting material maintained at an electrical potential [27].

Drift tubes (DTs) are used in the barrel region ($|\eta| < 1.2$). They consist of 4 cm-wide

tubes of gas with a positively-charged wire stretched along the centre. 4 layers of drift cells form a superlayer, and 2 or 3 superlayers form a drift chamber. The chambers are in turn grouped into stations; 4 layers of stations surround the CMS detector.

In the endcap regions, cathode strip chambers (CSCs) are used. Each gas-filled chamber consists of anode wires crossed with copper cathode strips. The CSCs are grouped into 4 stations in each endcap, and cover the region $0.9 < |\eta| < 2.4$.

The third type of muon detector, the resistive plate chambers (RPCs), reside in both the barrel and endcaps. They provide a parallel detection system to the DT's and CSC's, and cover $|\eta| < 1.6$. RPCs consist of two high-resistivity plastic plates, one held at positive voltage and the other held at negative voltage, with gas in between. They have a fast response, with better time resolution but coarser spatial resolution than DTs and CSCs [25].

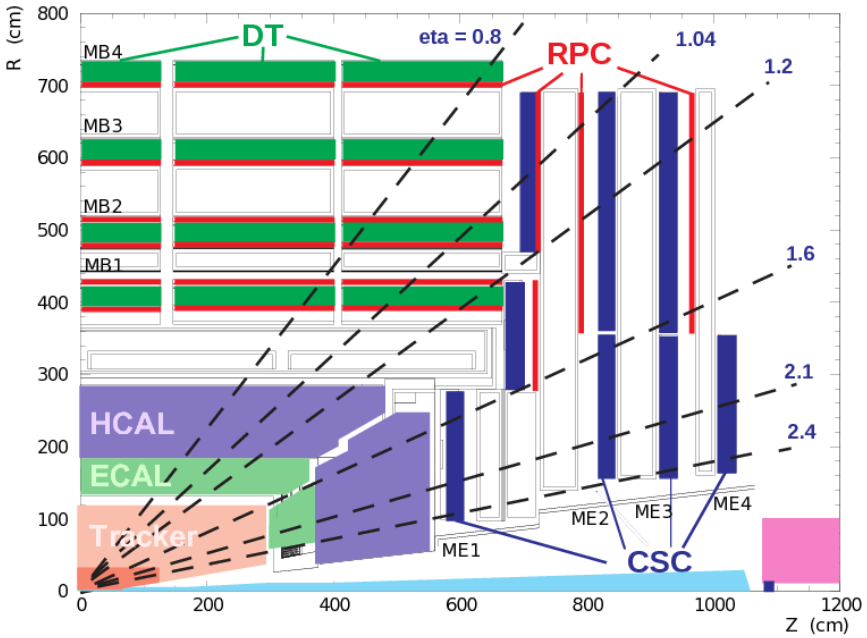


Figure 4.8: Schematic of the muon system, showing the three different types of gaseous detectors. Figure from [29].

4.3 The Trigger System

At a bunch crossing interval of 25 ns, 40 million bunch crossings take place per second, each with multiple collisions. The amount of data produced in these collisions is astronomical, too large to be read out and stored. Therefore, CMS has a trigger system which quickly decides, for each collision, whether it is an interesting event that should be stored, or a mundane one that should be discarded. The triggering takes place in two steps, called the Level-1 (L1) Trigger and the High-Level Trigger (HLT).

4.3.1 The Level-1 Trigger

The L1 trigger is the first line of defense against the huge onslaught of data from the collisions, and so has to be fast. It is made of custom-designed programmable electronics, such as field programmable gate arrays (FPGAs), application-specific integrated circuits (ASICs) and programmable memory lookup tables (LUTs). The process starts by generating trigger primitives, which are based on calorimeter energy deposits and hit patterns in the muon chambers (the tracker readout is too slow to be used in the L1 trigger). Successive stages of the trigger take in the trigger primitives and rank the particle candidates, feeding the result into the Global Trigger. The latter then makes the decision whether to accept or reject the event.

Part of the L1 Trigger resides on the detectors, and the rest is housed about 90 m away from the experimental cavern in an underground room. This trigger reduces the data output rate to about 100 kHz, introducing a latency of less than 4 μs. Its output is digitised and sent to the high-level trigger.

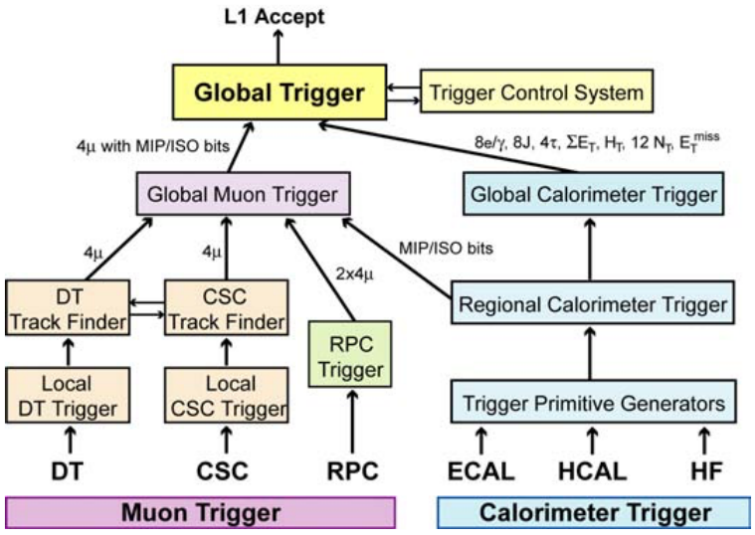


Figure 4.9: Architecture of the L1 Trigger. Figure from [25].

4.3.2 The High-Level Trigger

The HLT is a software-based trigger that does more complex calculations than the L1 Trigger, and runs on a large farm of processors. It reconstructs particles using some of the algorithms described in section 4.4, modified to optimise for speed. The HLT is able to use tracker information, allowing it to estimate momentum more precisely and thus identify particles more accurately. Reconstructing tracks and vertices is quite time-consuming, so the HLT may apply some simplifications – for example, disregarding information from the strip tracker, or keeping fewer track candidates at each step of the process than the full reconstruction algorithm.

4.4 Event Reconstruction

OK, so you have different bits of the detector lighting up at different times, and by different amounts. How do you then figure out, from that, what particles were actually passing through?

By a complicated process called particle reconstruction. We run algorithms that take information from disparate parts of the detector, and try to put it together to trace each particle's motion through the detector. By looking at things like the curvature of tracks and the amount of energy deposited in the calorimeters, we can get a value for the momentum and energy of the particles. The output of the algorithm is a list of particles in each event, with their identities and 4-momenta.

4.4.1 Tracks and Energy Clusters

The reconstruction algorithm starts by organising the hit data from the tracker and calorimeters. Track reconstruction is based on a Kalman Filtering algorithm. We start by generating seeds based on a few hits in the tracker – each seed needs to contain at least two hits in consecutive layers in the pixel detector. The track is then constructed layer-by-layer in the tracker. At each layer, the algorithm extrapolates the current track, predicting where tracker hits will occur in the next layer. These predictions are used to find compatible hits in the next layer, by doing a χ^2 comparison. Only tracks with at least eight hits in total, originate a few millimetres from the beam axis, and have a p_T

larger than 0.9 GeV are retained.

This process is repeated several times. At each iteration, hits from previously-found tracks are deleted, and the requirements on the number of hits and minimum p_T are also relaxed. This allows less obvious or lower-energy tracks to be found more easily. Later iterations also aim to reconstruct tracks with missing hits, or highly displaced tracks that only start in the strip detector [30].

Vertices, or the estimated points where collisions and decays occurred, are calculated by clustering tracks and using an algorithm called deterministic annealing [31]. Vertices can be near the beamline (indicating a collision), or displaced (indicating a decay). The primary vertex of interest is taken to be the one whose member particles have the highest total transverse momentum. The other vertices are called pileup vertices.

Meanwhile, energy deposits in the calorimeters are grouped together into clusters. First, cells with an energy larger than a given threshold, and larger than the energy of neighbouring cells, are identified as cluster seeds. We then grow topological clusters from these seeds, by sucking in neighbouring cells with an energy above twice the noise level. Once each topological cluster is grown, we cluster the cells within it using a Gaussian-mixture model, which models the energy deposits in the individual cells as being due to a certain number of Gaussian energy distributions, one for each seed. The means and variances of the Gaussians are fit using an expectation-maximisation algorithm, and their positions and energies then define the clusters.

4.4.2 Particle Flow

After forming the tracks and energy clusters, the next step is to link together elements from different detector components. The link algorithm compares pairs of elements that are near to each other in the η-ϕ plane. To link tracks to calorimeter clusters, the track is extrapolated into the calorimeter, and linked to clusters that it falls within. If more than one track is linked to the same ECAL cluster, or more than one HCAL cluster is linked to the same track, the link with the smallest distance is selected.

Each group of linked elements is then assigned a particle type, depending on which detector elements were involved. Muons are composed of tracks in the tracker and hits in the muon system. There are three types of muons: standalone muons are reconstructed by forming tracks in the muon chambers; global muons are reconstructed by matching muon chamber tracks to inner tracker tracks; and tracker muons are reconstructed by starting from tracker tracks and extrapolating them to the muon system.

Electrons tend to emit a large amount of energy in the tracker in the form of Bremsstrahlung radiation. When reconstructing electrons, this energy is gathered by forming a supercluster of clusters in the ECAL within a narrow window in η and extended window in ϕ of the electron trajectory. The extended ϕ window accounts for the electron's bending in the magnetic field. The tracks are then fitted with a Gaussian-sum filter (GSF) algorithm, which is more suited to electrons than the Kalman filter, and passed through a boosted decision tree classifier.

Isolated photons are reconstructed from ECAL superclusters with a transverse energy greater than 10 GeV, which are not linked to a GSF track.

Once muons, electrons and isolated photons have been reconstructed, they are removed

from the list of particle flow elements. The remaining ECAL clusters which are not linked to tracks are identified as nonisolated photons. HCAL clusters not linked with tracks are classified as neutral hadrons. The remaining HCAL clusters are then linked with tracks and possibly ECAL clusters. If the momentum of the track is roughly equivalent to the total energy of the clusters, the particle is identified as a charged hadron. If the track momentum is much smaller than the energy of the clusters, however, the particle flow object is classified as a photon, and possibly a neutral hadron, overlapping with a charged hadron [30].

4.4.3 Jets

The quarks and gluons produced in collisions are coloured, and cannot exist individually due to colour confinement. As they fly out from the collision point, they combine with quarks and antiquarks that were spontaneously created from the vacuum, forming hadrons. This process, called hadronisation, produces a collimated shower of particles called a jet.

Particle flow (PF) candidates are clustered together to form jets using the anti-k_T clustering algorithm [32]. This uses the distance metric

$$d_{ij} = \min\left(p_{Ti}^{-2}, p_{Tj}^{-2}\right) \frac{\Delta_{ij}^2}{R^2}\,, \tag{4.4}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is a measure of the angular distance between the two particles, with y being the rapidity. R is a cone size parameter that corresponds to the maximum radius of a jet produced using this algorithm, and is usually set to 0.4.

The algorithm runs iteratively. Distances between pairs of particles (or particles and

jets) are calculated, and the smallest distance determined. If this is less than the smallest p_{Ti}^{-2}, the two particles corresponding to the distance are combined into a jet. Otherwise, the particle or jet with the minimum p_{Ti}^{-2} is considered fixed, and removed from the list of objects.

Before jet clustering, all charged PF candidates whose tracks originate from pileup vertices are removed. Neutral candidates have no tracks, so their effect is estimated using the jet areas technique. Each jet's momentum is corrected by subtracting the median p_T density in the η-ϕ plane, multiplied by a factor called the effective jet area. This area is calibrated so that the pileup contribution is subtracted accurately.

4.4.4 b-Tagging

$t\bar{t}H$ processes where the Higgs decays into two b quarks have four b quarks in the final state, more than most background processes. This makes the presence of b jets a good indicator of such an event. To determine if a jet originated from a b quark, a b-tagging algorithm called the Combined Secondary Vertex (CSV) algorithm is used.

The CSV classifier takes as input several variables such as the number, position and mass of secondary vertices (vertices far away from the beam line), as well as the number of tracks and their characteristics. It then outputs a discriminator with a value between 0 and 1, with higher values indicating a higher probability of a jet being a b-jet. In the version of the algorithm used in Run II, an artificial neural network is used to output this discriminator.

4.4.5 Missing Transverse Energy

Some particles, such as the neutrinos that are relevant in leptonic $t\bar{t}$ processes, cannot be detected by CMS since they interact so weakly. Using the conservation of momentum, however, we can infer the existence of undetected particles. We do this by considering the negative vector sum of the transverse momenta of all reconstructed particles in the event:

$$\vec{p}_T^{\,\text{miss}} = -\sum_i \vec{p}_{Ti}\,. \tag{4.5}$$

Since the transverse momentum of all particles should sum to zero, a significant value of $\vec{p}_T^{\,\text{miss}}$ indicates the presence of undetected particles. (The total momentum in the beam direction cannot be assumed to be zero, because the protons accelerated by the LHC are composite particles, and individual pairs of colliding quarks may carry momentum in the z-direction.)

The magnitude of $\vec{p}_T^{\,\text{miss}}$ is called missing transverse energy, indicated by E_T^{miss} or MET.

CHAPTER 5

HOW TO FIND A NEEDLE IN A HAYSTACK: SEARCHING FOR $t\bar{t}H$ EVENTS

5.1 Monte Carlo Simulation

I noticed, when you were describing the CMS detector, that you didn't mention a section that detects Higgs bosons.

I didn't, because there isn't one! The Higgs, like many of the heavier particles produced by the LHC, decays too quickly to be directly detected. What we need to do instead is look at the particles that are detected in an event, and figure out if some of them could have come from a decaying Higgs boson.

How could you possibly know that?

The Standard Model predicts how the Higgs can decay (we call these decay channels). In this project, we're looking for events where the Higgs decays into a $b\bar{b}$ pair. What we need to do is to use the theory to predict what kinds of particles will be produced in different processes, and compare these with the particles we actually detect. We do this by running simulations using the Monte Carlo method, which allows us to generate large numbers of simulated events according to the probability distributions predicted by the Standard Model.

What goes on in a collider is messy business. Quark hadronisation and the interaction of particles with detector material makes it impossible to analytically calculate a probability distribution for the variables defining an event. Instead, simulations are done numerically, and take place in three stages.

In the first stage, we perform theoretical calculations, using Feynman diagrams to estimate the matrix elements of various physics processes. This produces a probability distribution over phase space, which we can randomly sample from; this sampling process is called the Monte Carlo (MC) method. The software package MadGraph is commonly used to produce the data in this stage.

The second stage simulates the showering and hadronisation process of quarks and gluons. This is often done by the Pythia software. To tune the shower parameters, generated events are compared to observed data.

The final stage involves a simulation of the detector response, with a package such as Geant4. This uses a detailed model of the geometry and material of the detector, including the magnetic field, and the interactions between the particles and the detector over different energy ranges. The MC sampling method is also used here to simulate stochastic effects. Finally, the software simulates the signals produced by the detector.

Lastly, the simulated detector signals go through the reconstruction process described in section 4.4, just like for observed data. The process of analysing MC events is then the same as that used to analyse data.

5.2 Processing Events

5.2.1 Preprocessing and Object Selection

In this analysis we look at single-leptonic $t\bar{t}H, H \to b\bar{b}$ events. We first select for such events by applying a trigger to detect a single lepton. Muon events must pass one of the two single muon triggers, `HLT_IsoMu24_v*` or `HLT_IsoTkMu24_v*`. Electron events must pass `HLT_Ele27_WPTight_Gsf_v*` [33].

When the MC simulation was generated, it was not known what the average amount of pileup in the data would be (since this number depends on the instantaneous luminosity, which varies over the course of a run). The pileup distribution in the MC set thus needs to be reweighted to match the data.

The leptons themselves must then undergo further selection. Muons are subject to the cuts $p_T > 26$ GeV and $|\eta| < 2.1$. They are also required to pass isolation requirements, which examine the particles in a cone of $\Delta R < 0.4$ around the muon. The isolation metric is

$$\mathrm{Iso}^{\mu} = \sum_{\Delta R<0.4} p_T^{\mathrm{CH}} + \max\left(0, \sum_{\Delta R<0.4} E_T^{\mathrm{NH}} + \sum_{\Delta R<0.4} E_T^{\mathrm{PH}} - \frac{1}{2}\sum_{\Delta R<0.4} p_T^{\mathrm{PU}}\right). \tag{5.1}$$

Here, CH indicates charged hadrons from the primary vertex, NH indicates neutral hadrons, PH photons, and PU charged hadrons from other vertices. The quantity Iso^{μ}/p_T is required to be less than 0.15. These requirements help to select muons that come from weak boson decays.

Electrons also undergo a similar selection process. They must fulfill the cuts $p_T > 30$ GeV, $|\eta| < 2.1$, and $\mathrm{Iso}^e/p_T < 0.06$. The electron isolation metric is

$$\mathrm{Iso}^e = \sum_{\Delta R<0.3} p_T^{\mathrm{CH}} + \max\left(0, \sum_{\Delta R<0.3} E_T^{\mathrm{NH}} + \sum_{\Delta R<0.3} E_T^{\mathrm{PH}} - \rho A\right), \tag{5.2}$$

where ρ is an energy density parameter and A is an effective area which is defined as a function of the electron η. Dedicated scale factors are then applied to MC events, in order to improve the lepton modelling's agreement with data. Finally, we only retain events with exactly one selected lepton.

Next, jets must be selected. After applying standard selection criteria [33], jets which are within a cone of $\Delta R < 0.4$ from a selected lepton are discarded. The remaining jets undergo calibration with jet energy correction scale factors, after which they are required to fulfill $p_T > 30$ GeV and $|\eta| < 2.4$.

b-jets are identified by applying the CSVv2 algorithm, and tagged if the discriminator exceeds 0.8484. Because the b-tagging efficiency is different between data and simulation, MC events have to be reweighted to account for this. This is done using the Tag-and-Probe method. We choose a sample with exactly two high-p_T leptons and two jets, at least one of which has been b-tagged, and apply cuts on the dilepton mass, missing transverse energy, and CSV of the tagged jet. These cuts are chosen to either select for dileptonic $\mathrm{t\bar{t}}$ events or Z+jets events. The former has 2 b-jets and can be used to measure the b-tagging efficiency, while the latter is used to measure the mistag efficiency. We then look at the CSV distribution of the second (probe) jet in separate bins of p_T and η. To mitigate contamination from jets of the wrong flavour, we split the MC samples into heavy and light flavour components, and subtract the non-relevant part from the data. By comparing the resulting CSV distribution between simulation and data, we can come

up with scale factors as a function of CSV, p_T and η (for both b-jets and light jets) with which we can reweight MC events.

A requirement that the missing transverse energy is at least 20 GeV is also imposed.

5.3 Discriminators and Distribution Shapes

So how do you measure the rate of $t\bar{t}H$ production? I expect it isn't as simple as just counting the number of times you see such a process...

You're right, it isn't. Because physics at the quantum scale is probabilistic, you can't know for sure whether any particular event is or isn't a $t\bar{t}H$ event. However, if you look at a very large number of events, you can estimate what percentage of them are the process you want.

How do you do that?

The trick lies in looking at distributions of certain variables that you measure in each event, and compare these with MC-generated collisions. The MC distributions show that the distributions of the signal process (the $t\bar{t}H$ events) have a different shape from the background processes, and by looking at the shape of the data distribution, we can figure out how much signal there is.

What kinds of variables do you use to make the distribution shapes?

Good question. That's what this section is about.

Statistically, several variables allow us to distinguish between $t\bar{t}H$ events and their largest background, $t\bar{t}$ + jets events [33]. These are:

- p_T of the jets in the event.
- The invariant masses of some subset of reconstructed objects in the event. A particular permutation of objects can form a subset; for example, we may take the mass of the dijet pair which is closest to the Higgs mass. We can also average over multiple permutations.
- The missing transverse energy.
- Angular separations (in η or ΔR) between pairs of objects. These are useful because they are relatively insensitive to the jet-energy scale uncertainty. As with the second point, these variables can be calculated for specific ranked permutations, or averaged over permutations.
- Event shape variables like sphericity.
- b-tag discriminant values for individual jets, or averaged over jets. These are useful for separating out the $t\bar{t}$ + jets background whose additional jets have been mis-tagged.

The most naive way to separate signal and background events would be to apply cuts on these variables. This doesn't work well, because signal and background events are usually not linearly separable. However, some of the preliminary cuts listed in sections 5.2.1 and 10.2.1 do give the signal a bit of an advantage.

Because of the low cross-section of $t\bar{t}H$ production (three orders of magnitude lower than the main $t\bar{t}$ + jets background), the events passing selection cuts are still dominated by background, whose statistical fluctuations would swamp any signal. We thus need to use more sophisticated methods to discriminate between signal and background. Naturally, we want to choose variables whose distribution shapes look as different as possible between signal and background.

In this analysis we use boosted decision trees (BDTs). BDTs belong to a special class of methods called multivariate analysis (MVA) methods. These allow us to use information from many variables at once to form our signal discriminator, a super-variable that provides higher discriminative power than using any variable individually.

5.3.1 Boosted Decision Trees

In a decision tree, each event is funneled through a series of yes/no decisions regarding the values of its variables. For instance, one node could assign an event to the left or right branch depending on whether its leading lepton p_T is greater or less than 40 GeV (Figure 5.1).

We construct a tree, choosing its cutoffs and splits, by training it on an MC set. Traditionally, we do this greedily for successive levels, optimising each node to give the best split. At each node, the value of the cutoff (the 40 GeV in this case) is chosen to maximise signal-background separation in the two daughter branches. We do this by calculating the signal purity p of each branch: assuming each event has a weight w_i, p is the total signal weight divided by the total weight of all the events in the branch. We

then calculate a measure of impurity, such as the Gini index:

$$\text{Gini} = p(1-p)\,. \tag{5.3}$$

Gini is low both for branches with mostly signal ($p \approx 1$) and branches with mostly background ($p \approx 0$). The best cutoff minimises the sum of Gini indices of the left and right branches.

Each branch is then split again using a different variable, with the cutoff chosen to maximise the change in Gini, and the process repeated until the whole tree is built. At each node, we pick the variable to use by choosing the one which yields the best separation [34]. We stop the process when some stopping criterion is reached (such as when we reach the maximum tree depth). The final branches are called leaves; leaves with $p \geq 0.5$ are signal leaves and the rest are background leaves [35] [36].

The classification accuracy of decision trees can be greatly improved by a technique called boosting, which uses an ensemble of different trees and averages their output. Specifically, we use the gradient boosting method. Instead of training the whole ensemble of trees at once, we start off with one tree, and add a new tree at each step. We calculate the quality of a tree (how well it does on the training set) by defining a loss function, a measure of how similar its outputs are to the true class of the training examples. At each step, we choose the parameters and weight of the new tree such that they will maximise the improvement to the loss function. This process has the effect of re-weighting events, such that events that had been misclassified in the previous step now have a higher weight [37].

In practice, choosing the optimal tree structures and set of input parameters is non-

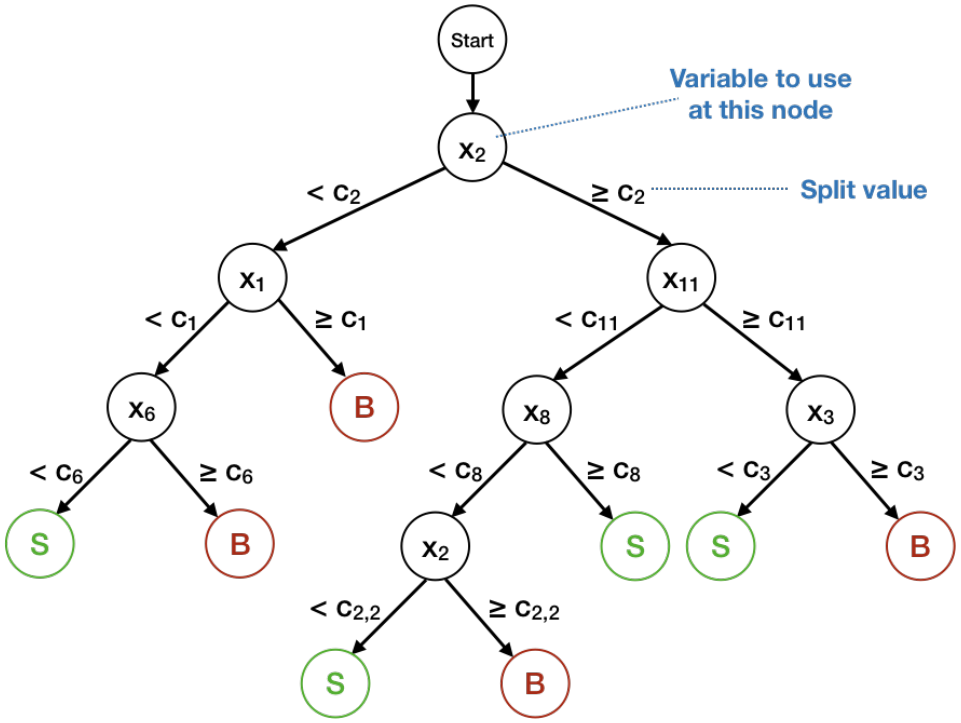


Figure 5.1: Decision tree.

trivial. One possible way to do this is to use the particle swarm algorithm (PSO) [38] [33]. PSO is an optimisation method inspired by the movements of individual birds in a flock, whose goal is to find the optimum location to roost. The movement of the individual birds (or trees, in this case) through the phase space of possibilities is affected by a mixture of randomness, their neighbours' motion, their memory of the best location they have so far visited, and their knowledge of the best location found by any member of the flock so far. The swarm of trees explores the phase space of tree parameters, while the input parameters are changed simultaneously, and the best configuration is chosen as the one which minimises the integral of the receiver-operator characteristic.

Decision trees are very sensitive to the specific set of training events they are given, and prone to overtraining, so that slightly different training samples can lead to drastically different trees. One way of mitigating this problem is shrinkage: at each step of the

boosting process, we multiply the output of the new tree by a factor called the shrinkage parameter, which lies between 0 and 1. Using a small shrinkage parameter, such that later trees are weighted successively less than earlier trees, helps to reduce overfitting.

Another way of preventing overfitting is to use a variant of the gradient boosting algorithm called stochastic gradient boosting. At each step, we fit the trees on a subsample of the training set, drawn randomly without replacement. The fraction of events used at each step is called the bagging fraction.

Pruning is another technique that prevents overfitting by reducing the size of trees. Branches that do not provide much power in classifying events are cut away.

For this analysis, the TMVA package in ROOT was used to construct and train BDTs. Half of the dataset was used for training, with the other half set aside for the analysis. To select input parameters and optimise tree structure, the particle swarm algorithm was used [33]. Table 5.1 shows the input variables used for the BDTs in two categories separated by b-tag multiplicity.

5.4 Setting a Limit on Signal Strength

Armed with our discriminator, we can now use its distributions to measure the amount of signal in the data. More formally, we write the expected total yield $E(n_i)$ in each histogram bin i as the sum of our expected signal and background yields s_i and b_i:

$$E(n_i, \theta) = \mu s_i(\theta) + b_i(\theta). \tag{5.4}$$

≥ 6 jets, 3 b-tags	≥ 6 jets, ≥ 4 b-tags
HT	Average ΔR (tag, tag)
M (lepton, closest tag)	b-tagging likelihood ratio
Average CSV (tags)	Average $\Delta\eta$ (jet, jet)
M_2 of min ΔR (tag, tag)	Best Higgs mass
H_0	M_2 of min ΔR (tag, tag)
H_1	$\sqrt{\Delta\eta(t^{\text{lep}}, bb) \times \Delta\eta(t^{\text{had}}, bb)}$
$(\sum p_T(\text{jet}))/(\sum E(\text{jet}))$	aplanarity
M_2 (tag, tag) closest to 125	H_0
Average $\Delta\eta$ (jet, jet)	H_3
	Average CSV (tags)

Table 5.1: Input variables used in the BDTs for the categories ≥ 6 jets, 3 b-tags and ≥ 6 jets, ≥ 4 b-tags. M and M_2 both indicate the invariant mass, HT is the sum of the magnitude of the transverse momentum of all jets, and H_0, H_1 and H_3 are Fox-Wolfram moments (measures of event shape).

The expected signal and background yields are the counts that end up in each bin after the MC-generated events have been subject to the cuts, reweighting, and so on that we defined for the analysis. We've introduced a parameter μ, called the signal strength parameter, given by

$$\mu = \frac{\sigma}{\sigma_{\text{SM}}}. \tag{5.5}$$

This is the parameter that we're trying to measure – the best-fit value of μ compares the $t\bar{t}H$ cross-section σ that is consistent with our observations to the cross-section σ_{SM} predicted by the Standard Model. θ represents a full suite of nuisance parameters, which are parameters that we allow to vary in the fit, but whose values we are not interested in. These parameters represent the systematic uncertainties which can affect distributions and yields.

We can then write a likelihood function for the observed event yields:

$$\mathcal{L}(\text{data}|\mu,\theta) = \text{Poisson}\left(\text{data}|\mu \cdot s(\theta) + b(\theta)\right) \cdot p(\tilde{\theta}|\theta). \tag{5.6}$$

Here, "data" represents the observed counts, while $p(\tilde{\theta}|\theta)$ is the probability distribution of the nuisance parameters (which I will describe later). Introducing the nuisance parameters into the fit reduces the impacts of the uncertainties, since the fit both constrains them and introduces correlations between different sources of uncertainties [7] [6]. The Poisson distribution is given by

$$\prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-\mu s_i - b_i}, \tag{5.7}$$

where the index i runs over the bins, and we see n_i events in bin i.

5.4.1 The Confidence Level Limit

The best-fit value of μ is then the one that maximises the likelihood function. However, we want more than a best-fit value of μ – we want a measure of how *certain* we are that our observations are (or are not) consistent with the Standard Model cross-section. We thus present our results as a 95% confidence level (C.L.) limit on the signal strength modifier: $\mu^{95\%CL}$.

A simple, Bayesian way of calculating the 95% C.L. upper limit would be to plot the likelihood in Equation 5.6 as a function of μ, and find the value of μ for which 95% of the area under the likelihood curve lies to the left of this μ:

$$\int_0^{\mu^{95\%CL}} \mathcal{L}(\mu)\, d\mu = 0.95\,. \tag{5.8}$$

This means that there is a 95% chance that the signal strength parameter μ is lower than $\mu^{95\%CL}$, or, in other words, that we can exclude any μ higher than this value with 95% confidence.

5.4.2 The Profile Likelihood Test Statistic

Instead of calculating the limit in this way, the LHC employs a frequentist approach by using a test statistic. While there are several possible ways to define this test statistic, we use one that is based on the profile likelihood ratio [39] [40]:

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})}\,, \qquad \text{with a constraint } 0 \leq \hat{\mu} \leq \mu\,. \tag{5.9}$$

Here, $\hat{\mu}$ and $\hat{\theta}$ are the values of μ and θ that maximise the likelihood function overall, while $\hat{\theta}_\mu$ are the values of the nuisance parameters that maximise the likelihood for a given value of μ [7]. Note that $\tilde{q}_\mu$ is a function of only μ, our parameter of interest (rather than also of the nuisance parameters, which we are not interested in). This allows us to define limits in terms of μ only [41]. Notice also that $\tilde{q}_\mu$ cannot be negative, since the denominator in the argument of the logarithm function is the maximum value of the likelihood for our data (so the numerator must be equal to or less than the denominator). $\tilde{q}_\mu$ gets more positive as μ gets further from $\hat{\mu}$.

Our next step is to figure out the distribution of the test statistic $\tilde{q}_\mu$. We do this by "throwing" toy Monte Carlo data, which we call pseudo-data. In general, this means that we get a count for each bin by randomly sampling from the Poisson distribution that defines that bin and the nuisance parameter distribution. In Higgs analyses, however, the values of the nuisance parameters are fixed to those that maximise the likelihood

function for the observed data ($\hat{\theta}_{\mu}^{obs}$ and $\hat{\theta}_{0}^{obs}$) when throwing pseudo-datasets. We throw two types of pseudo-data: one under the background-only hypothesis ($\mu = 0$ in the Poisson equation) and one under the signal hypothesis (using a particular signal strength μ). For each pseudo-dataset, we calculate $\tilde{q}_{\mu}$ (replacing the "data" in Equation 5.9 with the values from our pseudo-dataset). When calculating $\tilde{q}_{\mu}$, we allow the values of θ to float, performing a maximisation to find $\hat{\theta}$ and $\hat{\theta}_{\mu}$.

After throwing a whole bunch of toy datasets, we get two distributions for $\tilde{q}_{\mu}$ – one under the signal and another under the background hypothesis. Figure 5.2 shows these two distributions, for a background hypothesis for which $\mu = 0$, a signal hypothesis with signal strength $\mu = 1$, and $\tilde{q}_{\mu}$ calculated in both cases with μ set to 1. The background hypothesis distribution is flatter and spreads out more rightwards, which makes sense: $\tilde{q}_{\mu}$ is more positive when μ is further away from $\hat{\mu}$, and $\hat{\mu} \approx 0$ for the background hypothesis and ≈ 1 for the signal hypothesis.

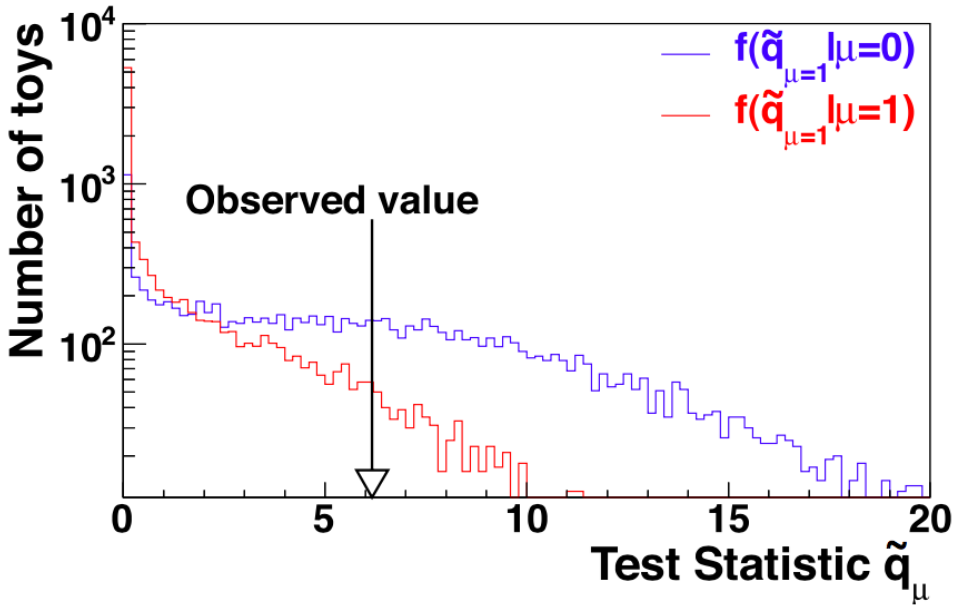


Figure 5.2: Distribution of test statistic with $\mu = 1$, for the signal + background hypothesis (red) and background-only hypothesis (blue). Figure from [39].

The "pile up" of events at $\tilde{q}_\mu = 0$ is due to the constraint on $\hat{\mu}$ that we set in Equation 5.9: that $0 \leq \hat{\mu} \leq \mu$. That is, we maximise the likelihood in the denominator, but only allow $\hat{\mu}$ to take on values less than μ. This means that for upward fluctuations of the pseudo-data, where we would have expected $\hat{\mu}$ to exceed μ, we instead have $\hat{\mu}$ getting "stuck" at the value of μ. $\tilde{q}_\mu$ then takes on a value of zero, instead of a positive value. When we calculate the area under the $\tilde{q}_\mu$ curve at its right-hand tail to obtain a p-value, these upward-fluctuating events won't contribute to the area, and won't increase the p-value. This is good, because upward-fluctuating events should not count against our signal hypothesis.

To obtain our limit, we need to calculate two p-values – one for each $\tilde{q}_\mu$ distribution. The p-value is the area under the curve to the right of the observed value of $\tilde{q}_\mu$. We call the p-value from the signal + background hypothesis curve p_μ, and the p-value from the background-only hypothesis curve $1 - p_b$. The confidence level $CL_s(\mu)$ is the ratio of these p-values:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b}. \tag{5.10}$$

The 95% C.L. upper limit for μ is then the value of μ for which $CL_s(\mu)$ is 5%.

This way of calculating $CL_s(\mu)$, by taking a ratio of two p-values, gives a conservative limit (that is, our actual confidence is higher than what it would indicate). Simply using the numerator p_μ would have given us a valid limit. However, taking the ratio makes our limit more robust against downward fluctuations of the data [39].

5.4.3 Nuisance Parameters

Equation 5.6 includes a term $p(\tilde{\theta}|\theta)$, where $\tilde{\theta}$ is the nominal (default) value of the nuisance parameter. To calculate this term, we express it in terms of pdfs $\rho(\theta|\tilde{\theta})$ using Bayes' theorem:

$$\rho(\theta|\tilde{\theta}) \sim p(\tilde{\theta}|\theta) \cdot \pi_\theta(\theta)\,. \tag{5.11}$$

We take the priors $\pi_\theta(\theta)$ to be flat.

A natural choice for the distribution of ρ would be a Gaussian. This works well for parameters that can be both positive and negative. However, it is not suitable for parameters that can only take on positive values, and which have large uncertainties that could exceed the value of the parameter itself. For these, the log-normal distribution is more appropriate:

$$\rho(\theta) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln\kappa)^2}\right)\frac{1}{\theta}\,. \tag{5.12}$$

For small uncertainties, the log-normal distribution approximates a Gaussian with standard deviation $\ln(\kappa)$.

When the number of events is small, the Normal distribution no longer approximates the Poisson distribution. For some uncertainties associated with the statistical uncertainties in control regions, the Gamma distribution is instead used:

$$\rho(n) = \frac{1}{\alpha}\frac{(n/\alpha)^N}{N!}\exp(-n/\alpha)\,. \tag{5.13}$$

Here, N is the number of events in MC or in a control sample in data, n is the number of events in the signal region, and α is the factor that we would multiply N by to get

the expected number of events in the signal region. The expected value of n is thus αN.

5.4.4 Expected Limits

The limit calculated above is the *observed* limit – the limit that we set using the physically observed data. But we would also like to know the SM expected limit – the limit on μ that we would *expect* to get (given the amount of data collected in a particular analysis and the uncertainties involved) if the SM cross-section is correct (that is, if $\mu = 1$). We also want to calculate the expected background limit, which is the limit we would expect if there were no $t\bar{t}H$ signal at all. These limits are important because even if $\mu = 1$, for example, we would still not be able to exclude values of μ down to 1 given the limited amount of data that has been collected and the uncertainties – the expected limit would still be higher than 1. The observed limit is thus compared to the expected limits to check for consistency with the SM or the background-only hypothesis.

The expected limit for the background-only hypothesis is calculated by generating many sets of $\mu = 0$ pseudo-data, and calculating $\mu^{95\%CL}$ for each set in the same way as we would for data. The median value among all the sets of pseudo-data is the median expected limit, and the 16% and 84% quantiles define the 1σ band, while the 2.5% and 97.5% quantiles define the 2σ band [39].

5.4.5 The Asymptotic Approximation

In practice, it is a bit of a pain to throw all these toy datasets needed to calculate the distribution of $\tilde{q}_\mu$ – we need one group of datasets for each value of μ to be tested, as well

as another group to calculate the median and 1σ and 2σ bands. Instead, we apply an asymptotic assumption – the assumption that the dataset has a large number of events N, such that terms of order $1/\sqrt{N}$ can be ignored. This approximation allows us to directly calculate formulae for the limits, using various statistical theorems.

In the asymptotic limit, the test statistic $\tilde{q}_\mu$ is equivalent to the test statistic q_μ, which is defined the same way as $\tilde{q}_\mu$, but without the requirement that $\hat{\mu} > 0$. Our upper limits are calculated in terms of this latter statistic.

First, we use a special artificial dataset called the Asimov dataset – defined as the dataset for which when we maximise the likelihood, the values of the parameters which yield the maximum are in fact the true parameter values [42]. In practice, the Asimov dataset is just the one whose counts in each bin are equal to the expected signal and background yields. In particular, we set $\mu = 0$, i.e. the Asimov dataset's counts are equal to the background-only yield for the nominal nuisance parameters. We calculate q_μ using the Asimov dataset, and call this value $q_{\mu,A}$.

The 95% C.L. upper limit for μ is then given by the solution to the equation

$$CL_s = \frac{1 - \Phi(\sqrt{q_\mu})}{\Phi(\sqrt{q_{\mu,A}} - \sqrt{q_\mu})} = 5\% \,, \tag{5.14}$$

where Φ is the cumulative distribution of the standard (zero mean and unit variance) Gaussian [39]. The $n\sigma$ band is then given by

$$\mu_n^{95\%CL} = \sigma \left(\Phi^{-1}(1 - \alpha\Phi(n)) + n \right) \,, \tag{5.15}$$

where $\sigma^2 = {}^{\mu^2}/_{q_{\mu,A}}$, and α is 5% in this case. To find the median expected limit on the

background hypothesis, we take $n = 0$ to give

$$\mu_{\text{median}}^{95\%CL} = \sigma\left(\Phi^{-1}(1 - 0.5\alpha)\right) = \sigma\Phi^{-1}(0.975)\,. \tag{5.16}$$

This approximation gives good results even away from the asymptotic limit, though if the number of events gets too small, it can be very over-optimistic [39].

CHAPTER 6

APPEALING TO THE PEOPLE: TECHNIQUES FOR EFFECTIVE SCIENCE COMMUNICATION

OK, OK, that's enough statistics for now! Here, why don't you ask me something about science communication?

Sure, I have plenty of questions. I was thinking of all the stuff you told me earlier, and it's really interesting food for thought. I always thought of science outreach as presenting facts and knowledge to the audience -- and the only challenge is to make things engaging and understandable, so that they don't fall asleep.

Unfortunately, facts and logic often aren't enough to convince somebody.

But they ought to! They're facts!

Perhaps they ought to, but the human brain isn't wired in this way. People build knowledge not by simply absorbing random facts that they encounter. Instead, every new block of knowledge has to be fit in with what's already there. If it doesn't fit, it often gets discarded.

Hmm, I suppose this has to do with how echo chambers form on the internet. People seek out stuff that confirms what they already believe, and ignore anything that challenges their world view.

Yep. This means that as a science communicator, you have to really think about your

audience – their background, their desires and interests and fears, their beliefs, what social groups they belong to. Only then can you design your communication in a way that will suit their needs.

6.1 The Audience is King

In a survey of 10 science communication experts, Bray et al. [43] found that they all agreed on one thing: that the audience comes first. They highlighted the importance of using empathy to understand an audience's needs and priorities, as well as to engage their imagination. They also emphasised that it is vital to build trust with one's audience, by being respectful and honest. This is easier said than done, especially when the science communicator is an expert in their field – they would need to remain humble and respectful towards the diverse opinions of a lay audience.

6.1.1 Audience Segmentation

Modern science communicators do not consider the public to be a homogenous body, neatly separated from scientists. Instead, they tend to use *publics* in the plural, to emphasise the different kinds of audiences one might encounter, and the fact that each may require a different approach. Demographic groups, such as age and education level, are obvious ways to divide up an audience. For example, an activity that is designed for children might not be suitable for adults; even among the latter group, something that attracts young professionals might not be appealing to retirees.

Audiences can also be divided according to their attitudes toward science. As an example, consider the subject of climate change. Leiserowitz et al. [44] propose that

the American public be split into six segments, named Global Warming's Six Americas. The segments are named the Alarmed, Concerned, Cautious, Disengaged, Doubtful and Dismissive groups. They differ not just in their beliefs, attitudes and behaviors regarding climate change, but also in their demographics, values and political opinions.

A communicator who wishes to encourage people to take action about climate change should handle each group differently [45]. The Alarmed and Concerned groups are more certain that human-caused global warming is happening, and tend to think about the issue and follow environmental news. In particular, the Alarmed tend to be more highly-educated than average, and lean left politically. Communication strategies for these groups should focus on explaining potential solutions to global warming, and encouraging them to take action, rather than presenting yet more evidence that it is happening. This audience is more willing to process difficult concepts, and so messages can be more complex and informative [45].

The middle two groups, the Cautious and Disengaged, do not think very much about the topic, nor hold strong opinions, nor follow environmental news closely. They also tend to be less educated than the population average. Communication aimed at these groups should not require too much effort to process, instead focusing on evoking human responses – using humour, visual imagery and narratives, and promoting social norms [45].

Most tricky to approach are the Doubtful and Dismissive groups, who hold negative attitudes towards climate change. The Dismissive, especially, tend to be certain of the view that human-caused global warming is not happening. They are more well-educated than average, and lean right politically. The one question that these groups would most like to pose to a climate scientist is how one can know that climate change is real – but they are also unlikely to seek out this information, preferring to avoid reading about

the topic. A communicator targeting these groups should therefore be indirect and non-confrontational [45].

6.1.2 How Communication Efforts Can Backfire

A lack of understanding of one's audience can cause communication efforts to have the opposite effect from what one intended! Hart and Nisbet [46] call this the "boomerang effect", and it has been observed in a diverse set of topics, including anti-smoking campaigns, appeals for donations, and calls to protect the environment. Their study showed that while describing the effects of climate change increased the level of support for climate mitagation policy among Democrats, it sometimes *decreased* support among Republicans.

One cause for the boomerang effect is cognitive bias – people tend to uncritically accept new information that supports their existing beliefs, while strongly resisting information that contradicts their beliefs. When presented with an unpalatable message, people might think of counter-arguments against it, thereby strengthening their certainty in their original position [45].

Another possible pitfall is that of underestimating the audience's intelligence. If a particular audience is inclined to carefully process information, it is important to present strong and logically-sound arguments to them. Attempting to convince them using weak arguments might actually *discourage* them from changing their behaviour [45].

6.1.3 Social Communities

The social groups that somebody belongs to can shape their sense of identity, and are an important source of pride and self-esteem [47]. This sense of identity and community could even affect someone's attitudes towards science. While scientific facts may seem objective, their implications could be seen as a threat to someone's strongly-held cultural values, causing them to react defensively.

Climate change is a classic example. It is a tricky topic in part because it has become so highly politicised. Conservative people may react negatively towards discussions of climate change, because of the common implication that the government should impose regulations on commercial activity – a threat to the value of individualism. To mitigate this effect, one might frame the conversation in a way that is more congruent with conservative values. For example, suggesting nuclear power, rather than government intervention, as a solution to climate change causes these groups to be more accepting of information on the topic [48].

Science communication efforts during the Ebola outbreak in West Africa illustrate a different way in which cultural considerations are important. Communicators had to address traditional practices that increased the risk of transmission, such as caring for the sick at home, burial practices, and seeking care from traditional healers. Low literacy rates meant that messages had to be mostly pictorial, but images had to be culturally-appropriate as well. The Centers for Disease Control and Prevention worked with local partners from each country to produce materials targeted to a specific community. Because outsiders tended to be viewed with mistrust, health officials enlisted the help of trusted community members, who in turn disseminated information to the rest of the population [49].

6.2 The Humanity of Science

However much audiences may differ, they are all human, after all. In general, therefore, it works well to appeal to human emotions in communication. This may be an unpalatable idea to scientists, who like to think of science as the objective pursuit of truth.

But science is also an inherently human endeavour. The passion that drives scientists, their frustration and disappointments, the challenges they must overcome, and the elation they finally feel upon solving a problem – all this could make for as riveting a read as any novel. To present the people behind the science, to tell a story with a scientist as the protagonist, is an oft-used technique which draws in the audience and makes them care about a subject which could not otherwise have held their attention.

A particularly well-executed example of this is the documentary film *Particle Fever*, which tells the story of the Higgs boson discovery. Along the way, it focuses on four or five scientists (both theorists and experimentalists), who talk in interviews about things like how they got into physics, their hopes for the accelerator, and their fear lest it not discover anything. It even shows them going about their daily lives, rowing or running or discussing things at blackboards. Importantly, the film does not shrink from portraying negative events – in particular, the accident that occurred just after the first start-up of the LHC. It succeeds in making the audience feel for the physicists whose life's work seemed to be in jeopardy. Many reviewers praise this emotional aspect of the movie, with some saying that by the end, they were wholeheartedly cheering along with the physicists when the Higgs boson was finally found [50]. Had the film merely contained a discussion of how the Higgs was discovered, and why it is so important to physics, it would be hard to imagine it inspiring such a fervour of feeling.

Scientists are not the only ones with a role to play in a story about science. In many topics, such as health and the environment, the ordinary people who are affected by an issue also make compelling characters. Telling their stories can evoke sympathy and draw in the audience. The audience, themselves, may also have a personal interest in a topic, which one can bring to light by explaining the applications and impact of a scientific discovery.

Other ways to capture attention include invoking the audience's sense of wonder (such as in *Animal Planet* documentaries). One might play up the mystery of a particular topic, presenting the journey of a discovery like a detective story, revealing the solution in a surprise twist at the end. Subjects such as astronomy or paleontology also have the potential to make the audience wonder about their place in the world [51].

Emphasising the humanness of science may sometimes take away from explaining the science itself – the technical concepts and details. But this is not necessarily a bad thing, depending on one's goals. New knowledge gained is often quickly forgotten, and one certainly cannot hope to teach difficult concepts within the framework of short activities. The feelings that are invoked, on the other hand, can last much longer. If the object is to excite enthusiasm for and interest in science, then appealing to emotions is certainly a good technique.

6.3 The Imperfection of Science

So there's something that I've always wondered about. I thought about it when you mentioned that scientists like to idealise science as this objective pursuit of truth. But actually, scientists know how

imperfect science is -- all the uncertainties and correlations that we have to deal with. I would have said that it's the *public* who think that science has all the answers, and is never supposed to be wrong.

Yes, there's certainly an idea in the public consciousness that scientists should know everything.

I think scientists themselves like to project that image!

Some of them, perhaps! But this wasn't always the case – Louis Pasteur, for instance, would conduct demonstrations of his experiments in public, so that they could be scrutinised by other scientists. People back in the day were more used to the idea of leading scientists "slugging it out" in public (16).

Oh, this reminds me of the Great Debate in 1920, which was an event publicly held at the Smithsonian. It was a debate between two opposing camps in astronomy -- the proponents of the Big Bang theory, and those who thought that the universe was expanding in a steady state.

When it comes to a less innocuous topic than the history of the universe, though, disagreement between scientists can be a very tricky thing to handle. As a journalist, it's important to present *both* sides of an argument, in order to be objective.

That seems pretty straightforward.

The tricky part comes when both viewpoints aren't equally held within the scientific community. If we return to the topic of climate change, for example – the majority of climate scientists agree that humans are causing climate change. But if you

present both views equally, you have the problem of false equivalence. That is, it sounds like both views are equally valid, when in reality one has much more support.

So what do you do?

You could stress that a certain viewpoint is not held by many experts (52). Or, you could sort of "shift the balance" by presenting a viewpoint that isn't quite in agreement with the majority consensus, but which also isn't completely out there. So you're putting the weighted midpoint of the opinion spectrum roughly in the right place.

That sort of makes sense. I suppose another tricky thing about science is that it's inherently uncertain... in some fields because we just can't make very good predictions yet (like in earthquake science), or because scientists are human and may sometimes make mistakes.

Yep. That's an issue because the public, and journalists, *like* certainty. And they like simplicity, too. They want a clear, unambiguous answer to a question – is eating chocolate healthy or not?

But of course science can't answer a question like that -- there are so many confounding factors!

Of course, but nobody wants to write a headline that says that eating one bar of dark chocolate per day over a month was correlated with decreased cholesterol levels in a 10-person sample group. Putting in caveat after caveat, as scientists like to do, is confusing – and also doesn't make for good clickbait!

But you have to state some caveats, or at least explain how the scientists reached a certain conclusion...

You have to find a balance. If you present science as perfect, people will have unrealistic expectations and there may be a backlash if the scientific establishment ever changes their mind about something. But if you focus too much on the uncertainty, people will start thinking that their opinion is as good as an expert's, since the experts aren't sure about anything anyway! Scientists tend to err on the side of emphasising uncertainty too much, because those are the areas that they spend most of their time researching and debating with their peers. But when it comes to public communication, you don't want to forget that there's lots of stuff that most of your colleagues agree on.

A good way to introduce this imperfect nature of science is to work it into the story you're telling. Say you're telling the story of a discovery from a scientist's point of view – you could show the starts and stops, the mistakes made, talk about how they followed the wrong path for a while.

Yeah, this reminds me -- I've found in my experience that when I show vulnerability, say if I admit that I find certain physics concepts difficult, or that I sometimes struggle with tying what I do to the more practical aspects of life -- it actually seems to make people feel better-disposed towards myself and what I'm studying.

6.4 Science Writing

So you've mentioned science journalism a few times in our conversation so far. Do you do a lot of writing yourself?

Some – I enjoy it.

Ah, maybe you could give me some tips then. Most of the writing I've had to do is for scientific papers, academic stuff -- and I know that's very different from writing for a lay audience. But despite being a scientist myself, I actually find it much more enjoyable to read a popular science article than a scientific paper. In fact, if I'm trying to learn about some topic in physics that I don't already know about, I first look for blogs that explain it, before resorting to academic papers!

The two kinds of writing are certainly very different! There's this quote I like by Quentin Cooper, a science journalist. He says:

"Science values detail, precision, the impersonal, the technical, the lasting, facts, numbers and being right. Journalism values brevity, approximation, the personal, the colloquial, the immediate, stories, words and being right now. There are going to be tensions."

Haha, right, we were talking about this in the previous sections. Scientists think that journalists don't explain the science with enough detail, while writers think that scientists don't consider

```
the human factor enough.
```

6.4.1 The Structure of Science Writing

One common way in which journalistic writing differs from scientific writing is in the order in which one presents things. In a scientific paper, one starts with giving an introductory background to the subject, citing work that has been done so far; then moving on to research methodology; then to the results of the current study, and finally one writes a short section on possible implications. A newspaper article does exactly the inverse! It follows the so-called "inverted pyramid structure" (Figure 6.1), starting with the most newsworthy information (the punchline – the results of the study, and any implications, often played-up). Next comes more detail (who did the research? how was it done?), and then the additional background information about the subject [53].

Figure 6.1: Inverted pyramid structure of news writing.

One can see why the inverted pyramid is favoured by journalists. People skimming through a newspaper want to get the most important ideas right away, before deciding

whether to read on for more detail. (There is also an interesting historical reason for this structure. When journalists would telephone or telegraph their stories in to the offices of their newspaper, there was a risk of the transmission being cut off. In that case, it was best to get the most important stuff through first. In addition, newspaper editors would often shorten articles to fit into the paper by simply cutting off sentences from the bottom – so those sentences couldn't contain any important information.)

Science writing that is not meant for newspapers (such as feature articles in science magazines) may not exactly follow the inverted pyramid structure. However, they must still capture the reader's attention right from the first paragraph. Often, this "hook" takes the form of the opening line of a story – a description of a scene, or the introduction of a character. The story then unfolds in the rest of the piece. An engaging hook is particularly important if the rest of the article will contain complicated information that is difficult to digest. It gives the audience a reason to put in the necessary effort.

As with any other form of writing, a piece should flow well, with one idea leading to the next. It will not do to simply list a bunch of facts. Unlike in academic writing, however, one has more freedom to use vivid language and detailed descriptions, which bring the story to life. A writer should not pepper a piece with jargon, of course, though introducing a few new words is acceptable as long as they are explained.

6.4.2 Explaining Scientific Concepts

So I get that it's important to have a narrative, and to focus on the human factor and all that. But you can't *just* do that, can you? I mean, then you might as well be writing a novel. You do have to

include some of the difficult scientific stuff.

Right, in many cases you do. How much difficult stuff vs. how much narrative you include will vary depending on your purpose and audience. If this were a feature story about a discovery, you might tell quite an extensive yarn about the people involved, even describing their backstory as children! But if you were writing a blog post that aims to explain some scientific concept, then you might spend more time on the technical stuff, since your readers are likely already interested in the subject and are truly curious to know the details.

So you have to, um, dumb things down by a different amount depending on your audience.

Well, I don't like the term "dumbing down", because that's being rather condescending towards your audience's intelligence! Besides, it takes quite a good grasp of a subject to be able to explain it to someone who doesn't know anything about it. If you can do that, you probably know it better than someone who can't explain it without using a lot of jargon.

That's right -- someone, I'm not quite sure who, once said that you don't really understand something unless you can explain it to your grandmother. Unless your grandmother happens to have a PhD in the subject, of course...

Before explaining anything, a writer should start by figuring out what the audience already knows about the topic. She can then guide them step-by-step through an explanation, making sure that each step follows logically from the last.

The art lies in knowing exactly the right amount of information to convey at each step, so that readers immediately understand what one is saying. When it's done right, it can be quite thrilling for a reader – to finally understand something that has puzzled them for a long time [54].

Of course, some amount of simplification is necessary when writing for a lay audience. It would hardly be appropriate to present the full mathematical derivations of general relativity when explaining what neutron stars are. Science writer Carl Zimmer describes what he calls the "science writer's dilemma" [55]:

"A good explanation achieves a happy medium between too little and too much [detail]".

I rather think this describes a balancing act more than a dilemma. If you give the audience too much detail (on the level of an academic paper), they will get bogged down and lose the big picture. If you just skim over the surface without going into detail, on the other hand, they will feel unsatisfied, and the story might even end up being uninteresting.

Leaving things out of a piece of writing is easier said than done. Having put a lot of time and effort into researching a topic, it can be quite painful not to include something! But as any artist will tell you, knowing what to leave out is just as important a skill as doing the writing itself.

As former Guardian science editor Tim Radford puts it, "Nobody has to read this crap" (56). The people who are reading your articles are probably doing so while standing in an overcrowded subway train during rush hour, squinting at their phone screen. At any minute, if things fail to capture their interest, they may just stop reading. So it's important to always write for the reader, and not for yourself!

CHAPTER 7

TOP RECONSTRUCTION BY KINEMATIC FITTING

Speaking of complicated concepts that are difficult to explain, pardon me for saying so, but it seems like what you have to deal with at the LHC is... a huge mess! You have all these particles coming out in sprays and jets, and lots of collisions you don't care about mixed in with the interesting ones, and detector uncertainties, and...

Indeed. You can almost say that the whole existential purpose of an experimental particle physicist is to find tricks and stratagems to wade through the mess, so that you can measure something.

And what tricks and stratagems do you work on?

It's something called kinematic fitting. The idea is to get better estimates of measured values like particle momenta, by adjusting them so that they fulfill certain constraints.

What kind of constraints?

For example, in a decay, the tracks of daughter particles need to originate from the same point. Or, the daughter particles' invariant mass must equal the parent's mass.

Ah, I think I see – so you're saying that the measured values of particle momenta and other kinematic quantities will not fulfill these constraints exactly, because of

the measurement uncertainties that we were talking about – detector resolutions and uncertainties associated with reconstruction...

Right, so the goal is to nudge the measured values such that they fulfill the constraints, and hope that this new estimate is better. Here, I'll first describe the general method of kinematic fitting, and then I'll tell you about the specific one that I use...

7.1 Traditional Kinematic Fitting

The discussion in this section is adapted from a combination of [57], [58] and [59]. We start by introducing a χ^2 value that measures the distance between the measured values of variables, and our hypothesised values. In a simple case where there are only two uncorrelated variables, x_1 and x_2, the χ^2 would be

$$\chi^2 = \frac{(x_1 - x_{10})^2}{\sigma_1^2} + \frac{(x_2 - x_{20})^2}{\sigma_2^2}\,, \tag{7.1}$$

where x_{10} and x_{20} are the measured values of x_1 and x_2 respectively, and σ_1 and σ_2 are their uncertainties. Here we take x_1 and x_2 to be the hypothesised values. The goal is to minimise this χ^2 while taking the constraints into account. (Without the constraints, the values of the variables that minimise χ^2 would obviously be their measured values – not very helpful.)

In the more general case where we have n correlated measured variables, the χ^2 needs to be defined using the $n \times n$ covariance matrix V (here we arrange the n variables x_i

into a column vector):

$$\chi^2 = \Delta\vec{x}^T\, V^{-1} \Delta\vec{x}\,, \tag{7.2}$$

where $\Delta\vec{x}$ is the difference between the measured and hypothesised values of $\vec{x}$. Suppose that we also have m constraints f_k, which are a function of the measured variables $\vec{x}$ and p unmeasured parameters $\vec{a}$:

$$f_k(a_1, ..., a_p, x_1, ..., x_n) = 0\,. \tag{7.3}$$

In order to minimise the χ^2 subject to these constraints, we use m Lagrange multipliers λ_k. The problem then becomes equivalent to finding the minimum of the function

$$\mathcal{L} = \Delta\vec{x}^T\, V^{-1} \Delta\vec{x} + 2\vec{\lambda}^T \vec{f}\,. \tag{7.4}$$

Here, $\vec{\lambda}$ and $\vec{f}$ are m-dimensional, while $\vec{x}$ is n-dimensional.

To find this function's minimum, we need to set its derivative to zero. This is easier if we expand each constraint to first order about the point $(\vec{a}_0, \vec{x}_0)$:

$$f_k(\vec{a}, \vec{x}) = \sum_{j=1}^{p} \Delta a_j \left.\frac{\partial f_k}{\partial a_j}\right|_{a_j = a_{j0}} + \sum_{i=1}^{n} \Delta x_i \left.\frac{\partial f_k}{\partial x_i}\right|_{x_i = x_{i0}} + f_k(\vec{a}_0, \vec{x}_0) \tag{7.5}$$

where a_{j0} and x_{i0} are the initial measured values of the parameters a_j and x_i respectively. If we then define the $m \times p$ matrix A and the $m \times n$ matrix X such that $A_{kj} = \frac{\partial f_k}{\partial a_j}$ and $X_{ki} = \frac{\partial f_k}{\partial x_i}$, Equation 7.5 can be written in matrix form,

$$\vec{f}(\vec{a}, \vec{x}) = A\Delta\vec{a} + X\Delta\vec{x} + \vec{f}(\vec{a}_0, \vec{x}_0) \tag{7.6}$$

where $\vec{f}$ is the m-dimensional vector of constraints. The thing we want to minimise is then

$$\mathcal{L} = \Delta\vec{x}^T V^{-1} \Delta\vec{x} + 2\vec{\lambda}^T \left(A\Delta\vec{a} + X\Delta\vec{x} + \vec{f}(\vec{a}_0, \vec{x}_0) \right) , \tag{7.7}$$

and setting its derivatives in $\vec{x}$, $\vec{a}$ and $\vec{\lambda}$ to zero gives us the equations

$$\begin{aligned} V^{-1}\Delta\vec{x} + X^T\vec{\lambda} &= 0 \\ A^T\vec{\lambda} &= 0 \\ X\Delta\vec{x} + A\Delta\vec{a} + \vec{f}(\vec{a}_0, \vec{x}_0) &= 0\,. \end{aligned} \tag{7.8}$$

These can also be written in matrix form as

$$\begin{pmatrix} V^{-1} & 0 & X^T \\ 0 & 0 & A^T \\ X & A & 0 \end{pmatrix} \begin{pmatrix} \Delta\vec{x} \\ \Delta\vec{a} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\vec{f}(\vec{a}_0, \vec{x}_0) \end{pmatrix} . \tag{7.9}$$

This equation can be easily solved for our desired hypothesised values of $\vec{x}$ and $\vec{a}$ by calculating the inverse of the leftmost matrix. These are our best-guess values of the variables, i.e. the ones that minimise the χ^2 given the constraints. Using Lagrange multipliers, we have reduced the minimisation process to a less computationally-intensive matrix inversion calculation.

If the constraints are linear, then Equation 7.5 is exact, that is, the partial derivative matrices A and X are constant. If the constraints are not linear, then we need to perform the above process many times, tweaking the variables a little each time, to converge on a solution.

After a bit more algebra, one can obtain the new covariance matrix of our n variables [57]:

$$V_{\text{new}} = V - VX^T(XVX^T)^{-1}XV + VX^T(XVX^T)^{-1}A(A^T(XVX^T)^{-1}A)^{-1}A^T(XVX^T)^{-1}XV\,. \tag{7.10}$$

This can be shown to have diagonal elements smaller than the original covariance matrix V [58]. Thus, the uncertainties of the variables decrease after kinematic fitting. We can think of the constraints as introducing more correlations between the variables, so that some part of their uncertainties (the diagonal terms in V) is "transferred to" the correlations (non-diagonal terms in the matrix).

7.2 Kinematic Fitting Using Ellipses

So that's it? You just invert a matrix?

Well, no. Remember that that only works for linear constraints. Many of the constraints that we have to deal with, like invariant mass constraints, aren't linear in the momentum variables.

So do you do the thing you mentioned before, where you repeat the matrix inversion multiple times with slightly different variables each time?

Not quite. We try to do it in a smarter way, by using some of the constraints to reduce the number of parameters we have to minimise.

This variation on the kinematic fitting method, which I'll call "kinematic fitting with ellipses" or "top reconstruction using kinematic fitting", applies to $t\bar{t}$ events produced in conjunction with some other particles [60]. This method combines the covariance matrix kinematic fitting method described in the previous section with some additional analytic manipulations for some of the constraints, to form a χ^2 for each event.

The method is particularly useful in leptonic $t\bar{t}$ events for coming up with best estimates for the neutrino momenta, which we cannot measure. Briefly, the process goes as follows: we apply kinematic constraints from the top decay to obtain a range of possible neutrino momenta, using the measured momenta of the other top quark daughters as well as the masses of the top and its daughters. For each neutrino momentum value in our allowed range, we vary the momenta of the other particles, subject to the constraint that the total transverse momentum of everything should add to zero. We then pick the configuration that yields the lowest χ^2.

For events that involve a dileptonically-decaying $t\bar{t}$ pair in conjunction with other particles, the final-state objects can be divided into three groups:

1. Measurable top daughters b and l^+, as well as anti-top daughters $\bar{b}$ and l^-

2. Unmeasurable top daughter ν and anti-top daughter $\bar{\nu}$

3. Non-top objects. We will assume that this only includes light jets, since we don't want to deal with any extra neutrinos.

For $t\bar{t}$ events where one or both W's decay hadronically, we will pretend that one of the daughters of each hadronically-decaying W is unmeasurable (at least for part of the reconstruction process). Thus the final-state particles are divided in the same way as above.

For the rest of this chapter, I will use 'top constituents' to refer to the daughters of both the top and anti-top quarks.

7.2.1 Kinematics of Top Decay

Consider the kinematic equations that apply to a top (or anti-top) quark which decays leptonically:

$$E_t = E_W + E_b \tag{7.11}$$

$$\mathbf{p}_t = \mathbf{p}_W + \mathbf{p}_b \text{ (3 equations, one per dimension)} \tag{7.12}$$

$$E_W = E_l + E_\nu \tag{7.13}$$

$$\mathbf{p}_W = \mathbf{p}_l + \mathbf{p}_\nu \text{ (3 equations, one per dimension)} \tag{7.14}$$

$$m_t^2 = E_t^2 - |\mathbf{p}_t|^2 \tag{7.15}$$

$$m_W^2 = E_W^2 - |\mathbf{p}_W|^2 \tag{7.16}$$

$$m_\nu^2 = E_\nu^2 - |\mathbf{p}_\nu|^2 \tag{7.17}$$

The unknown quantities in these equations are E_t, $\mathbf{p}_t$, E_W, $\mathbf{p}_W$, E_ν and $\mathbf{p}_\nu$ (everything else can be measured in the event, and we assume the masses are known). The 11 equations and 12 unknowns leave one degree of freedom for the unmeasurable neutrino momentum.

What does this one degree of freedom look like? Section 2 through 2.5 of Betchart et al. [61] lead us through a derivation which shows that these kinematic equations constrain the neutrino momentum to lie on an ellipse in 3D momentum space in the laboratory frame. To summarise their derivation, the kinematic constraints from the $t \to bW$ decay

(Equations 7.11, 7.12, 7.15 and 7.16) constrain the W momentum to an ellipsoid of revolution about the axis of the b-jet momentum vector. Meanwhile, the kinematic constraints from the $W \to l\nu$ decay (Equations 7.13, 7.14, 7.16 and 7.17) constrain the neutrino momentum to an ellipsoid of revolution about the lepton momentum vector. But by Equation 7.14, translating the $\mathbf{p}_W$ ellipsoid by the vector $-\mathbf{p}_l$ should give another surface of solutions for the neutrino. Our final solution for the neutrino momentum is thus the intersection of this translated ellipsoid and our first neutrino ellipsoid – the intersection is an ellipse. The one degree of freedom corresponds to the angle which parametrises the ellipse, and which defines a point thereon.

Note that this kinematic derivation was done by considering only the particles involved in one top's decay. The two tops in a $t\bar{t}$ event can thus be treated separately.

7.2.2 Momentum Conservation Constraint and MET

How do we decide what point on the neutrino ellipse to choose? Consider a semileptonic $t\bar{t}$ event, where we have calculated the ellipse for the top that decays leptonically. A natural way to determine the best point on the ellipse is to choose the point that has values of $p_{\nu x}$ and $p_{\nu y}$ that are closest to the missing transverse momentum, $\not{p}_x$ and $\not{p}_y$ (see Figure 7.1). We could define a χ^2

$$\chi^2_{\text{neutrino,semileptonic}} = \frac{\left(p_{\nu x,\,\text{hyp}} - \not{p}_x\right)^2}{\sigma^2_{\not{p}x}} + \frac{\left(p_{\nu y,\,\text{hyp}} - \not{p}_y\right)^2}{\sigma^2_{\not{p}y}}\,, \tag{7.18}$$

where $\sigma_{\not{p}x}$ is the resolution of $\not{p}_x$ and $p_{\nu x,\,\text{hyp}}$ is the hypothesised value of $p_{\nu x}$, that is, the one obtained by choosing a point on the neutrino momentum ellipse. (And similarly in

the y-dimension.) In the case where $\sigma_{\not p x} = \sigma_{\not p y}$, for example, the minimum value of this χ^2 would correspond to the point on the ellipse (after it has been projected onto the x-y plane) which is closest to the MET (Figure 7.1).

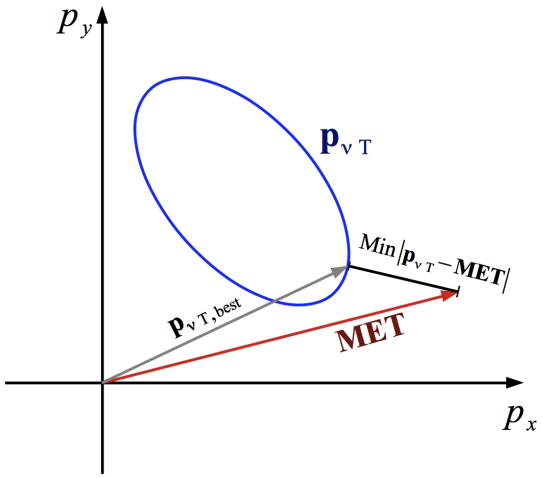


Figure 7.1: Best choice for a point on the $p_{\nu \mathrm{T}}$ ellipse (here shown projected onto the transverse x-y plane). The distance between this best point and the MET corresponds to the minimum χ^2.

What about a dileptonic event? Here we have two ellipses – one for each neutrino. We could impose the constraint that the neutrino transverse momenta add to the MET. In other words (working only with projections of the ellipses in the transverse x-y plane), if we were to take the $\bar{\nu}$ ellipse and flip it and translate it by $\not p_{\mathbf{T}}$ (that is, take the ellipse representing the quantity $\not p_{\mathbf{T}} - \mathbf{p}_{\bar{\nu}T}$), the points where this new ellipse intersects the ν ellipse are the solutions (for $\mathbf{p}_{\nu T}$) for which total transverse momentum is conserved for the event. $\mathbf{p}_{\bar{\nu}T}$ is then found by taking $\not p_{\mathbf{T}} - \mathbf{p}_{\nu T}$ (see Figure 7.2). Once the intersection points have been found, $p_{\nu z}$ and $p_{\bar{\nu} z}$ can be found by taking the corresponding points on the original 3D ellipses.

Because two ellipses on the same plane can intersect each other at either 2 points or 4 points (as shown in Figure 7.3; 1 or 3 points are theoretically possible but very rare),

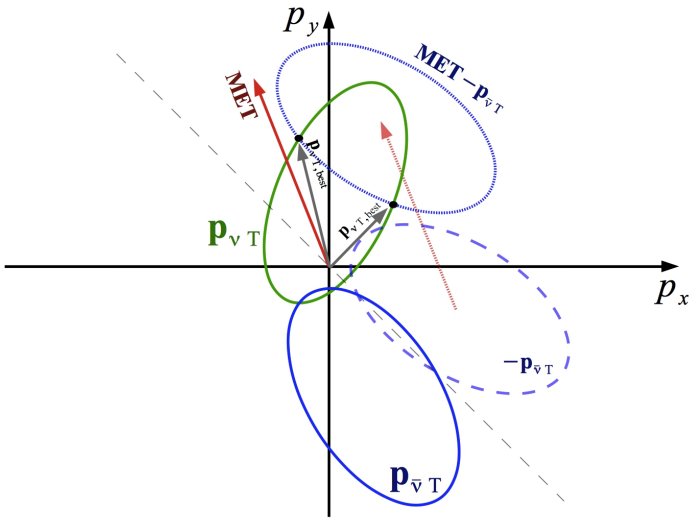


Figure 7.2: Take original $p_{\bar{\nu}\mathrm{T}}$ ellipse (solid blue), flip (dashed blue) and translate by MET (dotted blue). The two best choices for the value of $p_{\nu\mathrm{T}}$ are given by the intersection between the dotted blue ellipse and the original $p_{\nu\mathrm{T}}$ ellipse (green).

we can have either 2 or 4 solutions for best values of the neutrino momenta. It can also be the case that the ellipses don't intersect at all. In which case, our best value would correspond to choosing the point on each ellipse that is closest to the other ellipse, as shown in Figure 7.4. This would minimise a χ^2 that is defined by

$$\chi^2_{\nu\bar{\nu},\,\mathrm{dileptonic}} = \frac{\left((p_{\nu x,\,\mathrm{hyp}} + p_{\bar{\nu}x,\,\mathrm{hyp}}) - \not{p}_x\right)^2}{\sigma^2_{\not{p}x}} + \frac{\left((p_{\nu y,\,\mathrm{hyp}} + p_{\bar{\nu}y,\,\mathrm{hyp}}) - \not{p}_y\right)^2}{\sigma^2_{\not{p}y}}\,. \tag{7.19}$$

If the ellipses do intersect, this χ^2 would be zero at the intersection points.

7.2.3 Varying the Other Particles

While comparing the hypothesised neutrino momenta to the MET in this way is logical, the full potential of our reconstruction method could be realised if we consider the

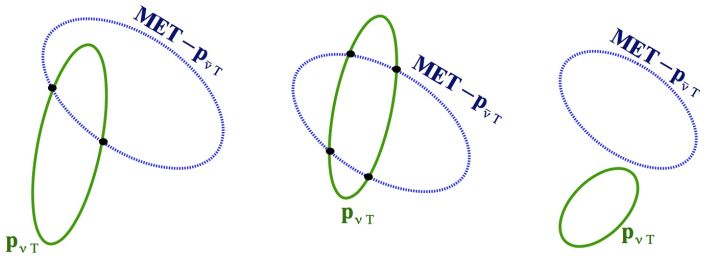


Figure 7.3: Three possible cases exist in the dileptonic case: two solutions (left), four solutions (centre) and no solutions (right).

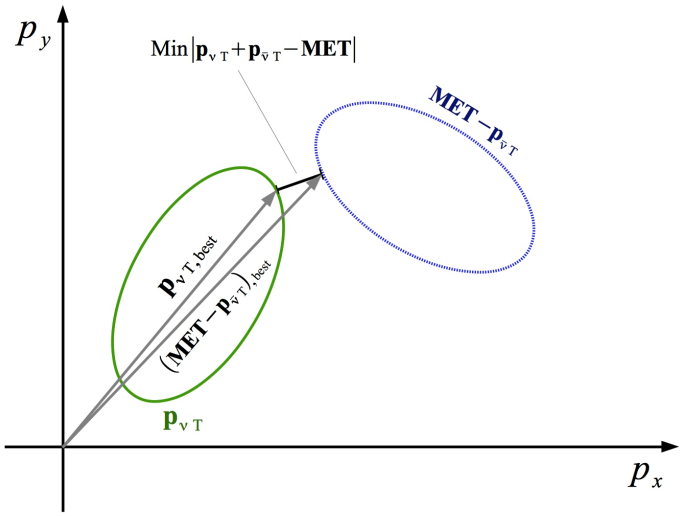


Figure 7.4: If there are no intersections, we could choose the point on each ellipse that is closest to the other ellipse.

fact that all the other particles (the non-neutrinos) in the event have measurement or reconstruction resolutions, and allow each of them to vary individually. Each object would then make a contribution to the total event χ^2. For instance, the b-quark's contribution would be

$$\chi^2_b = \frac{\delta^2_{bp_T}}{\sigma^2_{bp_T}} + \frac{\delta^2_{b\phi}}{\sigma^2_{b\phi}} + \frac{\delta^2_{b\eta}}{\sigma^2_{b\eta}}\,, \tag{7.20}$$

where $\delta_{bp_T} = p_{bT,\,\mathrm{hyp}} - p_{bT,\,\mathrm{measured}}$ etc. Here, the measured momentum is the one that has been measured and put through standard reconstruction methods, and the hypothesised

momentum is the one that has been varied.

Now, this wouldn't do us very much good if we were in the dileptonic case and the two neutrino ellipses intersect (after the first one is flipped and translated by MET) – in this case the neutrino contribution to χ^2 is zero, and if you jiggle everything else by a little bit, that would only shift the neutrino ellipses a little, so they would still intersect (albeit at slightly different points) and contribute zero χ^2. Meanwhile everything else has been jiggled and now gives a nonzero contribution to the χ^2. Thus you would get the lowest χ^2 (zero) by not jiggling anything at all.

If, however, the two ellipses do not intersect, we could try to jiggle everything else to make them move closer together and intersect. (Jiggling the measurable top constituents would change both the neutrino ellipses and the hypothesised MET; jiggling the particles that are not from the $t\bar{t}$ decay would change only the hypothesised MET. Either way, the distance between the ellipses changes. Here, the hypothesised MET is the one calculated from the sum of all hypothesised objects except the neutrinos.) If the decrease in $\chi^2_{\nu\bar{\nu},\,\text{dileptonic}}$ (due to the ellipses moving closer together) exceeds the increase in the χ^2 contribution from everything else (due to them being jiggled about), then we have a better configuration. We can thus choose the configuration that minimises the total χ^2.

A similar argument applies for the semileptonic case – if there is no point on the single neutrino's ellipse that exactly corresponds to the MET, we can jiggle everything else to change the ellipse shape and the MET, thus bringing the ellipse and the MET closer together.

7.2.4 Hadronic Top Decay

The derivation of section 7.2.1 does not assume a leptonically-decaying top – it applies equally well to a hadronically-decaying one. We can treat one of the W-daughters (let's call it Wq2, where q stands for (light) quark) as unmeasurable, calculating the ellipse that its momentum is constrained to lie on using the masses of the top, W, b and other W-daughter (Wq1) and the momenta of b and Wq1. We can then compare this ellipse to the measured value of the Wq2 momentum, so that Wq2's χ^2 contribution is

$$\chi^2_{\mathrm{Wq2}} = \frac{\delta^2_{\mathrm{Wq2}p_T}}{\sigma^2_{\mathrm{Wq2}p_T}} + \frac{\delta^2_{\mathrm{Wq2}\phi}}{\sigma^2_{\mathrm{Wq2}\phi}} + \frac{\delta^2_{\mathrm{Wq2}\eta}}{\sigma^2_{\mathrm{Wq2}\eta}}\,, \tag{7.21}$$

where $\delta_{\mathrm{Wq2}p_T} = p_{\mathrm{Wq2}T,\ \mathrm{hyp}} - p_{\mathrm{Wq2}T,\ \mathrm{measured}}$ etc.; here the hypothesised momentum is what we get from choosing a point on the Wq2 ellipse.

Since we can measure Wq2, why do we go to the trouble of calculating the ellipse? We can think of it as utilising the additional information given to us by the top decay kinematics, to find a configuration of object momenta that is better than the directly measured value. Whether a $t\bar{t}$ decay is hadronic, semi-leptonic or dileptonic, we would have two ellipses, one for each top, and each representing either a neutrino or a hadronic W daughter (from here on I will use "second W daughter", or "Wd2", as a general term referring to either the neutrino or Wq2). After calculating the ellipses, we can jiggle the other particles as discussed for the leptonic case, choosing as 'best configuration' that which produces the lowest total χ^2 for the event.

7.2.5 The χ^2 Minimisation Algorithm – Preliminary

The actual steps that we take to minimise the total event χ^2 are slightly different from those described above. For a start, because the effects of jiggling top constituents and non-top objects are different (the former also changes the ellipse shapes), we do these in separate steps. We first jiggle the top constituents, then use their jiggled values to calculate the ellipses, then jiggle the non-top objects. Also, the process of handling transverse momentum conservation is different, as we shall soon see. We use three nested minimisations thus:

1. Vary momenta of b, $\bar{b}$, W^+d1 and W^-d1.

    2. For a particular set of values of momenta of the particles in step 1, calculate the Wd2 ellipses for each top.

        3. Pick a particular point on each ellipse (defined by its parameterising angle θ), thus producing a hypothesised Wd2 momentum for each top. Vary the momenta of the non-top-constituents, **imposing the condition that total transverse hypothesised momentum of everything is zero**. (Here, 'everything' means all the particles: the hypothesised measurable top daughters, hypothesised Wd2's, and hypothesised non-top-constituents.) Calculate χ^2_{nonTop}, the non-top-constituent contribution to the chi-squared (for each non-top object this is of the form given in Eqn. 7.20; the total non-top contribution is the sum of each non-top object's contribution). Find the configuration of non-top object momenta that minimises χ^2_{nonTop}.

2R. Sweeping through possible pairs of θ values, calculate and minimise $\sum_{\text{tops}} \chi^2_{\text{Wd2}} + \text{Min}(\chi^2_{\text{nonTop}})$, where $\text{Min}(\chi^2_{\text{nonTop}})$ is the minimum non-top chi-squared found in step 3 for each pair of θ values. χ^2_{Wd2} is given by Eqn. 7.21 if the top is hadronic, and is zero if the top is leptonic (an explanation for this will follow in the coming paragraphs).

1R. Sweeping through values of momenta for the b, $\bar{b}$, W^+d1 and W^-d1 set, calculate and minimise $\sum_{\text{tops}} \chi^2_{\text{top}} + \text{Min}\left(\sum_{\text{tops}} \chi^2_{\text{Wd2}} + \text{Min}(\chi^2_{\text{nonTop}})\right)$, where $\chi^2_{\text{top}} = \chi^2_b + \chi^2_{\bar{b}} + \chi^2_{W^-\text{d1}} + \chi^2_{W^+\text{d1}}$ is each top's measurable-top-constituent contribution to the chi-squared and $\text{Min}\left(\sum_{\text{tops}} \chi^2_{\text{Wd2}} + \text{Min}(\chi^2_{\text{nonTop}})\right)$ is the minimum chi-squared value found in step 2R's minimisation. Each term in χ^2_{top} has the form of Eqn. 7.20.

Note that in this algorithm, because we impose the constraint of conservation of total transverse momentum when jiggling the particles, we do not need to treat MET as a separate object. Instead, after we choose a pair of points on the original Wd2 ellipses (without flipping or translating either of them to find intersections), we vary the non-top-constituent momenta in such a way as to give zero total transverse hypothesised momentum. This is mathematically simpler than the method described in sections 7.2.2 - 7.2.3, because we do not need to deal with the mathematics of moving two ellipses together until they intersect.

A further thing to note is that the two methods are not quite physically equivalent. In the new method, since we require that total transverse hypothesised momentum must be zero, the hypothesised MET is by definition equal to the total of the neutrino momenta we have selected. Thus we can no longer calculate the neutrinos' contribution to the χ^2 by comparing their hypothesised momenta to the MET; instead, neutrinos contribute nothing.

7.2.6 Non-Top-Constituents' Contribution to χ^2

The minimisations of step 1 and 2 are numerical minimisations (we scan through different values of momenta and pick the set that yields the lowest χ^2). Step 3, however, can be done in a smarter way. After choosing a pair of points on the ellipses and calculating the resulting Wd2's, we know the total hypothesised transverse momentum of the top constituents, $\sum_{\text{tops' daughters}} \mathbf{p}_{T,\text{ hyp}}$. To impose zero total transverse momentum, we need the total transverse momentum of the non-top constituents to cancel this, as well as momentum from any other particles that we're ignoring: $\sum_{\text{non-top}} \mathbf{p}_{T,\text{ hyp}} = -\sum_{\text{tops' daughters}} \mathbf{p}_{T,\text{ hyp}} - \sum_{\text{others}} \mathbf{p}_{T,\text{ measured}}$. The last term has the subscript *measured* instead of *hyp* because we're not jiggling these "other" particles, so their "hypothesised" momentum is just the same as their measured momentum.

This momentum conservation constraint entails, for a difference vector $\mathbf{d}$ that we now define,

$$\begin{aligned} \mathbf{d}_T &\equiv \sum_{\text{non-top}} \mathbf{p}_{T,\text{ hyp}} - \sum_{\text{non-top}} \mathbf{p}_{T,\text{ measured}} \\ &= -\sum_{\text{tops' daughters}} \mathbf{p}_{T,\text{ hyp}} - \sum_{\text{others}} \mathbf{p}_{T,\text{ measured}} - \sum_{\text{non-top}} \mathbf{p}_{T,\text{ measured}}\,. \end{aligned} \tag{7.22}$$

This is a known quantity since all three terms in the second line are known when we enter step 3. We then require

$$\sum_{\text{non-top}} \delta_{iT} = \mathbf{d}_T\,, \tag{7.23}$$

where $\delta_i = \mathbf{p}_{i,\text{hyp}} - \mathbf{p}_{i,\text{ measured}}$ for the i^{th} non-top object.

This is a linear constraint if we work in Cartesian coordinates, so we can handle the non-top-constituents in the same way as in the basic kinematic fitter method described

in section 7.1. That is, their covariance matrix is used to minimise their contribution to the χ^2, subject to the constraint in Equation 7.22 that their total transverse momentum must balance that of the other particles. This step of the minimisation thus simplifies to a single matrix inversion.

The measurement uncertainties for momenta are usually given in cylindrical coordinates because of detector topology, with σ_{p_T}, σ_ϕ and σ_η assumed to be independent. However, we would like to do the Lagrange multiplier minimisation in Cartesian coordinates, because the constraint is linear there. We therefore have to convert the covariance matrix for the non-top-constituents to Cartesian coordinates, as described in [60].

7.2.7 Varying Top and W Masses

Our reconstruction procedure can also take into account the uncertainties in top and W masses due to Breit Wigner widths, varying them and calculating their χ^2 contribution. The top mass contribution to χ^2 (and similarly for the W mass) is:

$$\chi^2_{m_t} = \left[\Phi^{-1}\left(\frac{1}{\pi}\tan^{-1}\left(\frac{\delta^2_{m_t} + 2\delta_{m_t}\cdot m_{t,\text{nominal}}}{\sigma_{m_t}\cdot m_{t,\text{nominal}}}\right) + \frac{1}{2}\right)\right]^2 , \tag{7.24}$$

where Φ^{-1} is the inverse of the cumulative distribution function of the lower tail of the standard Gaussian, $\delta_{mt} = m_{t,\text{ hyp}} - m_{t,\text{ nominal}}$, $m_{t,\text{ nominal}}$ is our known value of top mass, and σ_{m_t} is the resonance width.

7.2.8 Range of Top Mass for Which Valid Solutions Exist

There is one more step which we need to take. Recall from section 7.2.1 that each neutrino's ellipse is calculated by taking the intersection of two ellipsoids. What happens if the two ellipsoids don't intersect? It turns out that for each top, there is a variable which Betchart et al. [61] call Z^2 which acts as a flag for this. If $Z^2 > 0$, the ellipsoids intersect in an ellipse; if $Z^2 = 0$ they touch at a point; if $Z^2 < 0$ there is no intersection. Z^2 is a function of the top and W masses as well as of the b and Wd1 momenta.

We handle the cases for which the ellipsoids don't intersect by playing with the top mass. For a particular set of values of the momenta of b and Wd1 and of m_W for each top, we calculate the range of m_t for which $Z^2 > 0$. During the minimisation, we only vary m_t within this range. This ensures that we are always able to calculate the ellipses in step 2.

7.2.9 The χ^2 Minimisation Algorithm – Final

Our final minimisation algorithm thus looks like this:

1. Vary momenta of b, $\bar{b}$, W^+d1 and W^-d1, as well as the value of two m_W variables (one for each top).

    2a. Pick a particular set of values of momenta of the particles in step 1. For each top, find the range of m_t for which Z^2 is positive (the range can be different between the two tops).

        2b. For each top: for a particular m_t in the above range, calculate the

Wd2 ellipse.

3. Pick a particular pair of points on the two ellipses. Calculate and minimise χ^2_{nonTop}, imposing zero total transverse hypothesised momentum.

2R. Sweeping through possible pairs of θ values and values of m_t in the allowed range, calculate and minimise $\sum_{\mathrm{tops}} \chi^2_{\mathrm{mt}} + \sum_{\mathrm{tops}} \chi^2_{\mathrm{Wd2}} + \mathrm{Min}(\chi^2_{\mathrm{nonTop}})$.

1R. Sweeping through values of momenta for the b, $\bar{b}$, W^+d1, W^-d1 and m_W set, calculate and minimise $\sum_{\mathrm{tops}} \chi^2_{\mathrm{top}} + \sum_{\mathrm{tops}} \chi^2_{\mathrm{mW}} + \mathrm{Min}\left(\sum_{\mathrm{tops}} \chi^2_{\mathrm{mt}} + \sum_{\mathrm{tops}} \chi^2_{\mathrm{Wd2}} + \mathrm{Min}(\chi^2_{\mathrm{nonTop}})\right)$.

This method differs from the traditional kinematic fitting method described in section 7.1, in that it treats kinematic constraints in two different ways. The kinematics of the top and W decay are used to put a constraint on Wd2 in the form of restricting it to lie on an ellipse. On the other hand, the constraint of total transverse momentum conservation enters into step 3 via Lagrange multipliers.

CHAPTER 8

OPTIMISING THE TOP RECONSTRUCTION FITTER

That... didn't sound complicated at all.

For the benefit of our readers, let me just put in that that was said in a sarcastic tone.

... Right. It seems like it would probably be complicated to implement as well – especially the three nested minimisations.

Yes, it was pretty tricky to get the minimisation to converge.

What do you mean – converge? Don't you just try out all possible values of all the momenta, and find the configuration that gives the lowest χ^2?

Oh no no, you can't do that -- there are 18 parameters to be minimised, and sweeping through all of that phase space would take forever. Instead you have to use a gradient-based method... Here, before I explain what kinds of problems I had with the minimisation, I'd better talk a bit about how the minimisation is done in the first place...

8.1 Minuit Minimisation Algorithms

To do the minimisation we use the MINUIT package, which is available in ROOT. MINUIT was first developed in the 1970s, and specialises in minimising functions with

many parameters, as well as difficult problems which may require guidance to find a solution [62]. In this section I will describe the two main minimisation algorithms used by MINUIT: `Migrad` and `Simplex`.

8.1.1 `Migrad`

The most naive way to find the minimum of a function is to do a grid search – to sample the function at all possible points within a given range, down to a certain resolution. While this method is stable (it always converges), it takes far too long when we are dealing with a high-dimensional space of many parameters. Even modifications to grid search, such as first sampling with a coarse resolution and then "zooming in" to the best area and searching again with a finer resolution, are often too slow [63].

Instead, many minimisation algorithms make some assumptions about the function to be minimised – that they are fairly continuous and smooth, and don't vary too wildly over small distances. This is a reasonable assumption for functions representing physical quantities. An algorithm can then make use of the gradient of the function to move in the direction of a minimum. While this method is much faster than grid search, it isn't guaranteed to converge on the global minimum; instead, we must be content with a local minimum.

Newton's Method

To see how `Migrad` uses a function's derivatives to find a minimum, let's first consider a one-dimensional quadratic function. Such a function is defined by three parameters (the intercept, coefficient of the linear term, and coefficient of the squared term). We

could also write this function as a Taylor series expanded about a point x_0:

$$\begin{aligned} F(x) &= F(x_0) + \left.\frac{dF}{dx}\right|_{x_0} (x - x_0) + \frac{1}{2} \left.\frac{d^2F}{dx^2}\right|_{x_0} (x - x_0)^2 && (8.1) \\ &= F(x_0) + g(x - x_0) + \frac{1}{2} G(x - x_0)^2 \,, && (8.2) \end{aligned}$$

where in the second line we have relabelled the derivatives. For our quadratic function, this expression is exact (i.e. there is no $O((x - x_0)^3)$ term). If we calculate the value of the function, as well as the first and second-order derivatives, at our initial point x_0, then the function is fully defined, and we immediately know where its minimum is: at $x_m = x_0 - g/G$ [63].

Instead of analytically calculating the derivatives, we usually estimate them by evaluating the function a small distance d away from x_0:

$$\left.\frac{dF}{dx}\right|_{x_0} \approx \frac{F(x_0 + d) - F(x_0 - d)}{2d} \tag{8.3}$$

and

$$\left.\frac{d^2F}{dx^2}\right|_{x_0} \approx \frac{F(x_0 + d) + F(x_0 - d) - 2F(x_0)}{d^2} \,. \tag{8.4}$$

In the more general case of multiple dimensions, the Taylor expansion is instead

$$F(\mathbf{x}) = F(\mathbf{x}_0) + \mathbf{g}^T(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \mathbf{G} (\mathbf{x} - \mathbf{x}_0) \,, \tag{8.5}$$

where $\mathbf{g}$ is the gradient vector, and $\mathbf{G}$ is the second derivative matrix (called the Hessian). The minimum of this function is at $\mathbf{x}_m = \mathbf{x}_0 - \mathbf{V}\mathbf{g}$, where $\mathbf{V} = \mathbf{G}^{-1}$ is the inverse of the second derivative matrix, called the covariance matrix.

Of course, real-world problems aren't usually quadratic. For non-quadratic functions, the above Taylor expansion still holds; but this time, it is not exact – there is a nonzero $O((\mathbf{x} - \mathbf{x}_0)^3)$ term. This means that when we take a step $-\mathbf{Vg}$ from our initial point $\mathbf{x}_0$, we would still not be on the exact minimum (but we'd hopefully be closer to it). We would then have to repeat the process, taking another step using the values of $\mathbf{V}$ and $\mathbf{g}$ calculated at the new point, to get even closer to the minimum. This minimisation technique is called Newton's method [63].

Positive-Definiteness of the Hessian

There is one obvious problem with this method, which we can easily see when considering the one-dimensional case: what if G is negative? Then we would be dealing with a concave function (at least in the region that we are currently in), and if we were to take a step of $-g/G$, we would end up near a *maximum* point. In the multi-dimensional case, a concave function corresponds to the case where $\mathbf{G}$ is not positive-definite. Thus, it only makes sense to take a step $-\mathbf{Vg}$ if $\mathbf{G}$ (or equivalently $\mathbf{V}$) is positive-definite.

Well, what do we do if it isn't? We can create a positive-definite matrix that is as "close" as possible to $\mathbf{G}$, by adding a term:

$$\mathbf{G}_{\text{forced posdef}} = \mathbf{G} + \lambda \mathbf{I}\,. \tag{8.6}$$

Here, λ is a positive scalar that is larger than the absolute value of the most negative eigenvalue of $\mathbf{G}$. We then take a step using the matrix $\mathbf{V}_{\text{forced posdef}}$ that is the inverse of this new $\mathbf{G}_{\text{forced posdef}}$.

There is yet another improvement that we can make to this method. Instead of

simply taking a step $-\mathbf{Vg}$, we perform a line search in the direction of $\mathbf{x}_0 - \mathbf{Vg}$ to find the minimum point along this line. We then go to this point for the next step [63].

The Variable Metric Method

A quadratic function's Hessian $\mathbf{G}$ is constant throughout phase space, while a non-quadratic function has a varying Hessian. This sounds familiar – it's rather like the metric tensor of general relativity, which is constant in Euclidean space but varies in non-Euclidean space. It turns out that the Hessian transforms just like a covariant tensor, so we can use it to construct quantities that are invariant under coordinate transformations.

The quantity

$$\Delta s^2 = \Delta \mathbf{x}^T \mathbf{G} \Delta \mathbf{x} \tag{8.7}$$

acts like an invariant distance measure. When our function to be minimised is a χ^2 function, Δs has a straightforward meaning – it's just the number of "standard deviations" that $\mathbf{x} + \Delta \mathbf{x}$ is away from $\mathbf{x}$.

Another very useful invariant quantity is

$$\frac{\rho}{2} = \frac{1}{2}\, \mathbf{g}^T \mathbf{V} \mathbf{g}\,, \tag{8.8}$$

which is the difference in function value between the current point (where $\mathbf{V}$ and $\mathbf{g}$ are calculated) and the minimum value of a quadratic function with this $\mathbf{V}$ and $\mathbf{g}$. We call this the *expected distance to the minimum*, or EDM, and `Migrad` uses this as a convergence criterion [63]. That is, if our EDM falls below a certain predefined value, we

say that the fit has converged and we have found the minimum.

One problem with using the second derivative matrix in our minimisation procedure is that it is very expensive to calculate – the number of terms is the square of the number of parameters in our minimisation. Instead of re-calculating it at every iteration, we instead assume that it is slowly-varying enough that we can estimate its value at the current step by applying a correction to its value from the previous step. The correction would depend on the values of $\mathbf{x}$ and $\mathbf{g}$ at the current step. There are various formulae for calculating this correction, one of which is Davidon's rank-two formula [63]:

$$\mathbf{V_1} = \mathbf{V_0} + \frac{\delta\delta^T}{\delta^T\gamma} - \frac{\mathbf{V}_0\gamma\gamma^T\mathbf{V}_0}{\gamma^T\mathbf{V}_0\gamma}\,, \tag{8.9}$$

where $\delta = \mathbf{x}_1 - \mathbf{x}_0$ and $\gamma = \mathbf{g}_1 - \mathbf{g}_0$. If we start off with a positive-definite $\mathbf{V}$, then provided that a line search was performed along the direction $\mathbf{x}_0 - \mathbf{V}\mathbf{g}$ at each iteration, the updated $\mathbf{V}$'s will always remain positive-definite.

This is called the variable metric method, and is `Migrad`'s basic algorithm: calculate the Hessian, force it to be positive-definite if necessary, and update the Hessian at each iteration using an update formula.

8.1.2 `Simplex`

While gradient-based minimisation algorithms are efficient, they are not as stable as could be desired, and sometimes do not converge. It is thus sometimes useful to use stepping algorithms, which perform a scan of the phase space. `Simplex` is an intelligent stepping method that uses information from the current step to decide what direction to

step in next.

In n-dimensional space, a simplex is a figure with $n+1$ vertices (the smallest number of vertices necessary to define an n-dimensional body). In 2D space it is a triangle, and in 3D space a tetrahedron. The minimisation algorithm is so-named because it evaluates the value of the function at $n+1$ points, updating one of the points at each step.

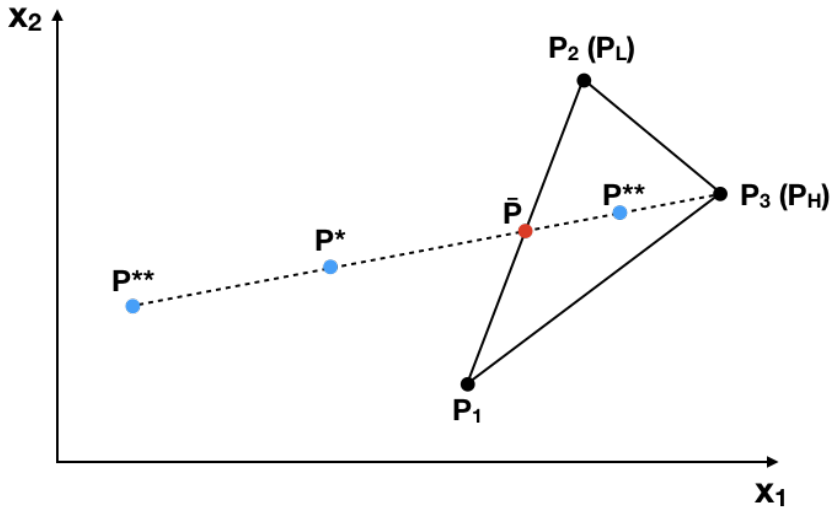


Figure 8.1: A simplex in 2 dimensions, showing the original three points P_1, P_2 and P_3, as well as the new points to try, P^* and P^{**}. Figure adapted from [63].

Here's how it works: starting with $n+1$ initial points, perhaps chosen randomly, we find the maximum and minimum points (which we'll call P_H and P_L respectively). Now we find the centre-of-mass point of all the points except P_H, and call this $\bar{P}$. We then try several new points to see if they are better than P_H. First we reflect P_H about $\bar{P}$ to get a new point P^*: $P^* = \bar{P} + (\bar{P} - P_H)$. If $F(P^*) < F(P_L)$, we go a bit further and try $P^{**} = \bar{P} + 2(\bar{P} - P_H)$. If $F(P^*) > F(P_H)$, however, we go in the opposite direction and try $P^{**} = \bar{P} - \frac{1}{2}(\bar{P} - P_H)$. We then take the lowest of these three points, and use it to replace P_H for the next step. If none of them are better than P_H, however, we form a new simplex about P_L, with dimensions reduced by a factor of 2 [63].

`Simplex` is designed to take as large steps as possible, and is efficient because it searches in a sensible direction (from the highest point to the average of the lowest points). As a convergence criterion, we can use the difference $F(P_H) - F(P_L)$. However, this convergence criterion isn't quite as trustworthy as that of the `Migrad` method [62].

8.2 Optimising the Top Reconstruction Fitter

Ok, point taken – finding the minimum of a function isn't quite as simple as I thought. Pun unintended. So what do you do if your fit doesn't converge?

To optimise the top reconstruction method and improve its convergence rate, I ran it on 1000 MADGRAPH-generated single-leptonic $t\bar{t}H$, $H \to b\bar{b}$ events. To simulate the effect of measurement uncertainties, the generated events were smeared, as described in Section 10.1.

The reconstruction algorithm suffers from two challenges. First is the three nested minimisations, two of which are numerical minimisations using MINUIT – a somewhat unusual configuration. With 14 parameters in the outer loop and 4 parameters in the second loop (which I'll call the inner loop), attaining convergence can be tricky. The second challenge is that the minimisation process must be run once for each event. Unlike problems where we do a single fit over all events, we do not have as much control over the process, and cannot "baby" the fit to get it to work.

8.2.1 Valid Solutions and Top Mass Range

After the majority of the bugs had been ironed out of the code and it seemed to be working correctly, its initial run (using the `Migrad` optimiser) attained a convergence rate of about 50%. Recall that, in order to ensure that our two ellipsoids intersect to give a solution, we restrict the top mass to the range in which $Z^2 > 0$. Originally, this range was calculated analytically by solving for roots of the Z^2 expression, using equations from [61].

However, the analytic top mass range calculator was buggy, often turning up ranges that didn't make physical sense (they were very far away from the nominal value) or that were overly restrictive (exactly at the nominal value). Since the equations are complicated and difficult to debug, I decided to remove the top mass range calculator altogether. Instead, I handled areas of phase space with no solution by introducing a flag that triggered whenever $Z^2 < 0$, which artificially set the value of χ^2 to one that was slightly larger than at the previous step. This led to an improved convergence rate of about 70 %.

MINUIT allows the user to control the optimisation process by running the minimisation multiple times, each time choosing a different set of parameters to fix or allow to vary. To further close the convergence gap, I attempted to guide MINUIT towards a solution by first fixing all parameters except the neutrino ellipse's θ parameter, perform a minimisation over θ, then fixing θ and releasing everything else, and then finally performing a third minimisation with all parameters free to vary. The idea was that since θ is the only parameter for which we have no measurement (so that its initial value is random, rather than measured), we would first allow it to vary into a reasonable region of phase space, before minimising over the other parameters. This produced a positive

effect – a convergence rate of 76 %. The events that failed showed a mixture of error code 3 (EDM did not attain the target value) and error code 2 (invalid Hessian).

To further isolate the problem, I ran a single minimisation with all particles fixed except the neutrino ellipse θ. With this configuration, 949 out of the 1000 events converged. Of the failures, 49 were due to an invalid Hessian. A similar picture presented itself when I allowed both the neutrino θ and Wq2 θ to vary: 950 events converged, with 49 failing due to an invalid Hessian. Crucially, all the events that failed with an invalid Hessian had wandered into a region of phase space with $Z^2 < 0$.

The problem, then, was that introducing a flag to artificially output a value of χ^2 whenever Z^2 dropped below zero had probably led to weird discontinuities in the function, thus producing problems when calculating derivatives.

Evidently, it was still necessary to calculate an allowed top mass range in advance. Instead of doing it analytically, as before, I implemented a crude scan over m_t, which steps through a range of 60 standard deviations with a step size of 0.1 standard deviations. At each step, the value of Z^2 is calculated, and finally the range of top mass for which $Z^2 > 0$ is returned (with the assumption that this range is contiguous, i.e. there are no two disjoint regions satisfying this criterion). The scan is done at the inner minimisation step, after the values of the outer minimisation parameters have been chosen, since the value of Z^2 depends on these parameters. After a few additional bug fixes, the convergence rate was now 76% when the minimisation was run in one pass with all parameters free to vary.

8.2.2 `Minimize`: `Migrad` and `Simplex` Hand-in-Hand

Since `Migrad` is a gradient-based method, it is quite dependent on the initial values of the parameters to be minimised. Smearing (or uncertainties and detector resolutions in the case of real data) might have moved the event into some unfriendly region of phase space from which `Migrad` cannot find a good minimum. In such a situation, a stepping method like `Simplex` might work better.

MINUIT offers a minimisation algorithm called `Minimize`, which combines `Migrad` and `Simplex`. First, a minimisation is done using `Migrad`. If this converges, that's the end of the story. If it does not converge, then the minimisation is done using `Simplex`. If `Simplex` finds a solution, `Migrad` is run again, using the solution that `Simplex` found as its initial parameter values. This allows `Simplex` to pick a good region of phase space to feed to `Migrad`, increasing the latter's chance of success. The final solution is then the solution returned by the second `Migrad` run [62].

In some cases, `Migrad` fails a second time – then the final solution returned is the one found by `Simplex`. If `Simplex` also fails, then the minimisation fails.

Using the `Minimize` algorithm produces a convergence rate of 84 % – a significant improvement.

8.2.3 Adjusting the Target EDM Value

So far, the target EDM had been set at 0.001 for both the inner and outer minimisers. Of the 16% of events that failed after the previously-mentioned improvements, most did so because the outer minimiser did not reach the target EDM. However, about two-thirds

of these actually attained EDM's that were pretty small – on the order of 0.01. It seemed that it might help if the target EDM were increased.

With a target EDM of 0.02 for the outer loop (and the inner loop's target remaining at 0.001), 92 % of events converged. Loosening the EDM target in this way does not seem to worsen the quality of the fit for the events that would have converged even with the lower target – these events still end up with low EDM's, on the order of what they had attained before.

It is important to keep the inner minimiser's target EDM relatively low. If it is increased by an order of magnitude, the outer minimisation starts to fail (even though the inner minimisation succeeds). It seems that the inner minimisation needs to return a fairly precise estimate of the inner loop parameters, in order for the outer minimisation to converge.

It is instructive to check if loosening the EDM target affects the "quality" of the best-fit parameters returned by the fitter. As a measure of result quality, consider the quantity

$$q = |p_{\text{fit}} - p_{\text{gen}}| - |p_{\text{measured}} - p_{\text{gen}}| \,. \tag{8.10}$$

Here, p is some kinematic quantity, which could be a component of momentum, or mass or energy. The subscript *gen* indicates the generated (ground-truth) value, *measured* refers here to the smeared value which we feed into our kinematic fitter, and *fit* is the value returned by the fitter. q compares whether the fitted result or original un-fitted quantity is closer to the true value. A negative value of q is good – this means that our fit has produced a better estimate of p than before the fit.

Figure 8.2 shows a plot of q against the final outer EDM value of events that converged,

for two kinematic quantities: the neutrino η and the mass of the leptonic top. The latter is calculated by taking the invariant mass of the neutrino, lepton, and daughter b quark. For the smeared case, which simulates directly measured data, we take the x- and y-components of the neutrino momentum to be equal to the MET (since there is only one neutrino in the single-leptonic case). The z-component of neutrino momentum is unknown, and is taken to be zero when calculating m_t and η_ν. For the generated case, $p_{\nu z}$ is known and used; for the fitted case, $p_{\nu z}$ has been estimated and is likewise used.

η_ν and $m_{t,\text{leptonic}}$ are two of the quantities that one would expect to be most improved by the fitter, since both depend on $p_{\nu z}$. The plots of q against final outer EDM show that the quality of events which converge with higher EDM does not seem much different from those which converge with lower EDM.

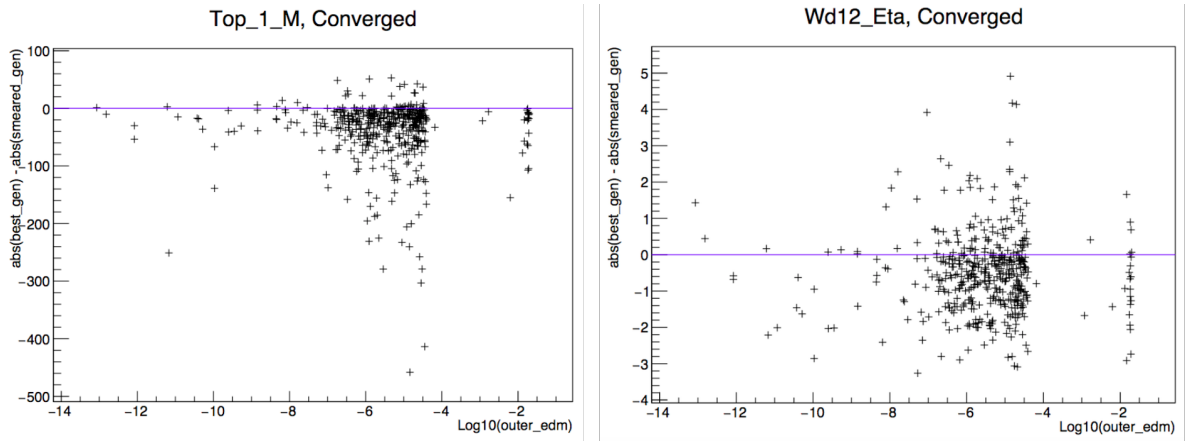


Figure 8.2: Result quality $q = |p_{\text{fit}} - p_{\text{gen}}| - |p_{\text{measured}} - p_{\text{gen}}|$ for η_ν and $m_{t,\text{leptonic}}$, for MADGRAPH-generated events whose fit converged. Top_1 is the leptonically-decaying top, and Wd12 is the neutrino.

8.2.4 Bug in Minuit

In the process of optimising the kinematic fitter, I also discovered a bug in MINUIT that was introduced between versions 6.02.00 and 6.06.00 of ROOT. The bug occurred at the code segment which was supposed to force the Hessian to be positive-definite. The variable containing the new Hessian was not declared in the correct place, and would go out of scope before the old Hessian could be updated. This caused the minimiser to be stuck in an infinite loop of checking for and forcing positive-definiteness.

I submitted a bug report to the ROOT development team, and the issue has now been fixed.

CHAPTER 9

TREMBLINGS AND WARBLINGS: A BLOG ON THE SCIENCE OF MUSIC AND SPEECH

"We have also sound-houses, where we practise and demonstrate all sounds, and their generation. . . . We represent small sounds as great and deep; likewise great sounds extenuate and sharp; we make divers tremblings and warblings of sounds, which in their original are entire. . . . We have also divers strange and artificial echoes, reflecting the voice many times, and as it were tossing it: and some that give back the voice louder than it came, some shriller, and some deeper; yea, some rendering the voice differing in the letters or articulate sound from that they receive."

- Francis Bacon, *The New Atlantis* (1627)

Wait, Francis Bacon wrote that? He talked about transforming sounds to make them higher or lower? And look, there's this line about adding *reverb* artificially!

Cool, isn't it? It's like he predicted the advent of electronic music, 300 years in advance.

That guy had some imagination.

Yeah. I stumbled across this work because I was trying to think of a name for my blog, and one of my friends suggested taking a phrase from this passage.

`You have a blog?`

Yep, it's about the science of music and speech. It's called *Tremblings and Warblings* – a name which I lifted from Bacon's passage. I thought it was fitting.

`What sort of stuff do you write about?`

Oh, anything that catches my fancy. How musical instruments work, the physics of music and sound, obscure historical technology, modern speech technology... I first became interested in the physics of music when I TA-ed a class on it, so I got the idea to start a blog about it. I write about basic concepts, and also about interesting stuff that I come across.

9.1 Blog Audience

Tremblings and Warblings (at `www.tremblingsandwarblings.com`) is an explanatory-style science blog, created using WordPress. Its goal is to explain concepts behind the science of music and speech, in a way that is engaging and easy to read.

As such, it is aimed at a specific audience – one that is already interested in science, music and sound. They find the topic cool, and want to learn more about it out of simple curiosity. They would be interested in knowing how things work to some level of detail, but do not necessarily have much background knowledge on the subject.

This is a more specialised audience than that of a more general publication such as a newspaper, or even some science magazines. Aiming at such a targeted audience is

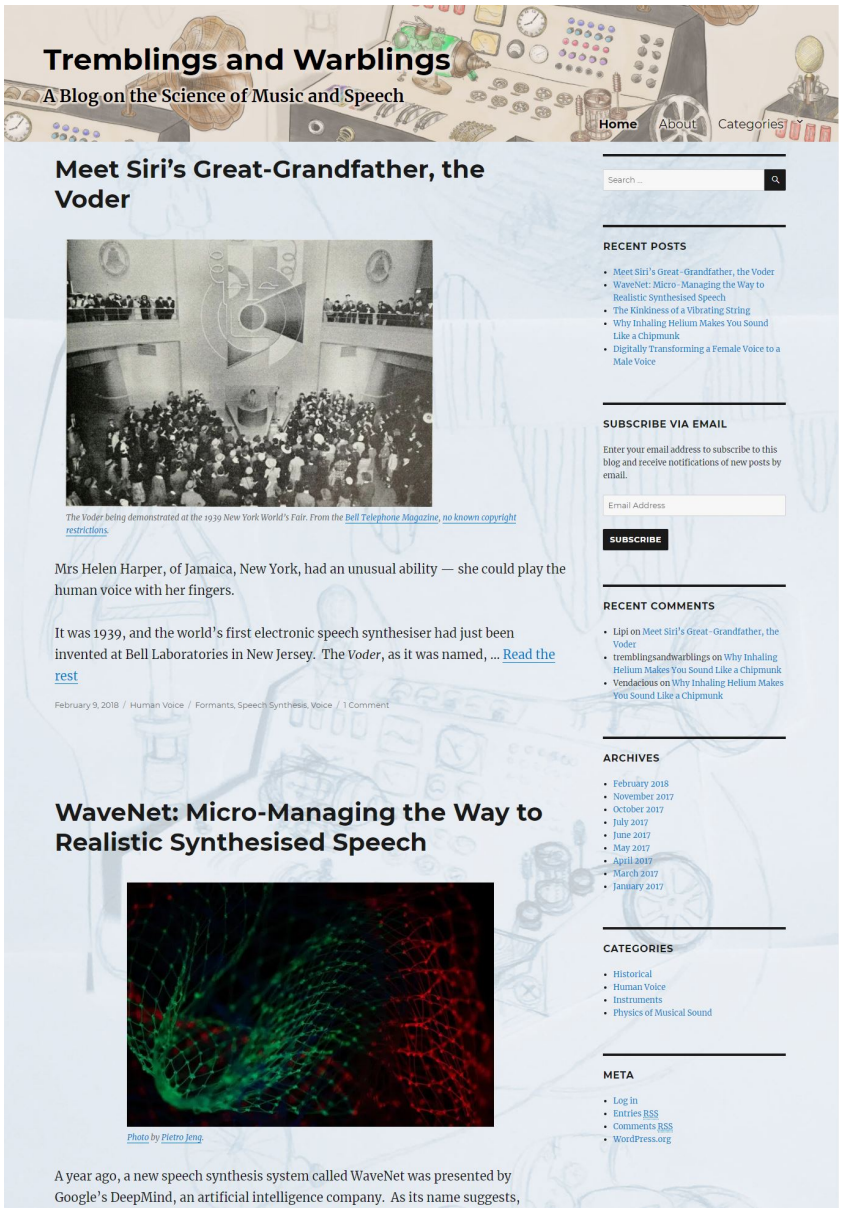


Figure 9.1: Blog homepage.

fair, however – professional science writers often make similar assumptions about their readers. They expect them to be intelligent, curious, and interested in science, though many may not have a background in it. Readers of science writing also often work in technical fields, and can belong to any age group [64].

As the audience for my blog grows and more of them interact via the comments section, I would get a better idea of their interests and background, and would then be able to adjust my writing accordingly.

9.2 Special Considerations for Online and Blog Writing

I guess that's the advantage of writing online -- you can target a very specific audience, because people have a lot of choice and control over what they read on the internet. But does online writing differ in other ways from more traditional forms of science writing?

Yep. The style can be quite different, for instance. You usually use a more informal and conversational tone, which makes the writer seem more accessible to the reader. Also, writers often use shorter sentences and paragraphs, because people are easily put off when confronted with a wall of text.

I expect the articles themselves are shorter as well?

Well, they used to be. It used to be conventional wisdom that online writing should be shorter and less detailed than print writing – that it should come in easily digestible bites, because people have short attention spans and can click away any time. But that's changed – it's now quite common to find really long blog posts and online feature articles that go into a lot of detail about a topic.

I guess that makes sense in a way -- you don't have any space constraints when writing online.

Yeah. Science blogs tend to be longer than other kinds of blogs on average, as well. Another expectation that's changed is how frequently you should post. People used to say that a blogger should post regularly, at least every week – any less frequent and you may lose readers who forget about your blog. But that's not so feasible when you're writing long, detailed posts. Besides, with RSS feeds and mailing lists, readers can set it up so that they are notified whenever there's a new post, so it's less vital to post often.

There's also a strong in-the-now feel to blogs, isn't there? I mean, blogs used to be sort of public diaries where people talk about what's happening in their lives at the moment.

Right, they have to be more relevant to current affairs than, say, an informative science website. But this also depends on the type of blog. If you're writing an explanatory post, or a how-to-do-something type of post, it may be more difficult to explain why you're writing it now rather than yesterday. Though of course it's good if you *can* link it to something that's currently relevant.

Oh, one other thing that surprised me when I first heard about it – the way you write *titles* is actually quite different between online and print writing! You actually have to be more direct when writing titles for online articles, because that makes them easier to find on search engines (65). There's less cryptic wordplay than in print articles. For example, my blog titles are usually pretty much a one-line summary of the post, like:

Why Inhaling Helium Makes You Sound Like a Chipmunk

Sometimes I can't resist using a more imaginative title, but in that case, I'll put in a colon followed by a subtitle that actually contains the article's keywords. Like this:

The Voice in the Soot: Humanity's Earliest Known Recording

9.3 Unique Aspects of Tremblings and Warblings

Given the deluge of information that an internet user is faced with every day, a blog needs to have a unique identity to avoid getting lost in the crowd. While good writing can draw in readers over time, I wanted to make my blog stand out in other ways as well.

9.3.1 Media, Sound Demos and Animations

A digital medium like a website is perfectly suited for explaining the science of sound, since the user can interactively play sounds and videos. I include as many sound demos and animations as possible in a post, to allow readers to see and hear phenomena for themselves.

Plots and Diagrams Based on Real Data

Some of the plots that one can find in textbooks and online resources about the science of sound are idealised or hand-drawn images, similar to the spectrum plots shown in Figure 9.2.

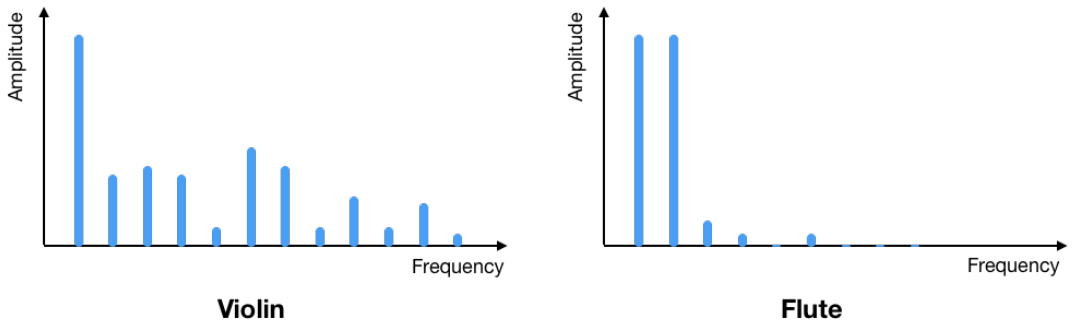


Figure 9.2: Idealised spectrum plots, similar to those found in [66].

While such plots can be very useful for illustrating concepts, I wanted to lend an extra level of realism to my demos by using plots which were generated from real data. Figure 9.3 shows an example: spectrum plots which I obtained by applying a Fourier transform to recorded violin and flute notes.

These plots are not perfect – notice the small peaks that are visible between the evenly-spaced harmonic peaks. However, they are more representative of "real life". Using plots generated from real sounds has a further advantage – I always include an audio clip of the corresponding sound, either with the image or nearby in the text, so that readers can compare what they see with what they hear.

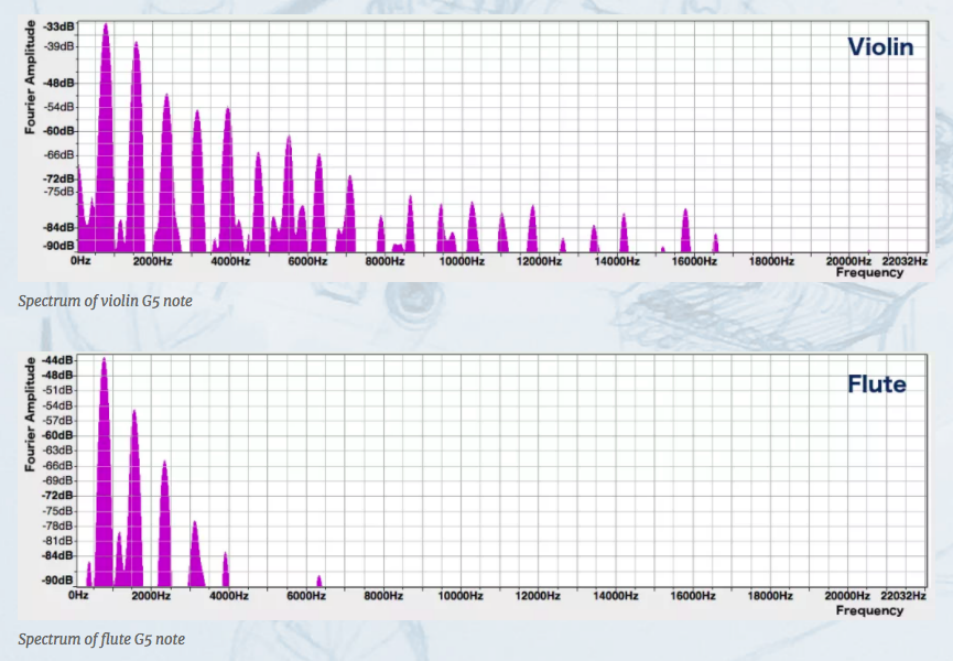


Figure 9.3: Spectrum plots of real violin and flute notes. Screen capture from one of my blog posts.

Self-Coded Sound Demos and Animations

Quite apart from analysing recorded musical sounds, I also created many demos from scratch. Some of these were manipulations of sound – an example is presented in the post *Digitally Transforming a Female Voice into a Male Voice*, where I transformed a recording of a woman speaking into one that sounded like a man. These demos allow me to use interesting effects to illustrate important concepts, such as that of formants, which are characteristics of the spectrum that allow us to distinguish between different sounds.

For some posts, I also created animations, with and without sound. The post *The Kinkiness of a Vibrating String* is a notable example – here, I made animations that show how sinusoidal standing waves on a string add to form a travelling kink that goes up and down the string (Figure 9.4). To achieve this effect, it was important to get the

amplitudes and phases of the component waves exactly right. Because I could not find these numbers anywhere, I calculated them from first principles using the Fourier series equations.

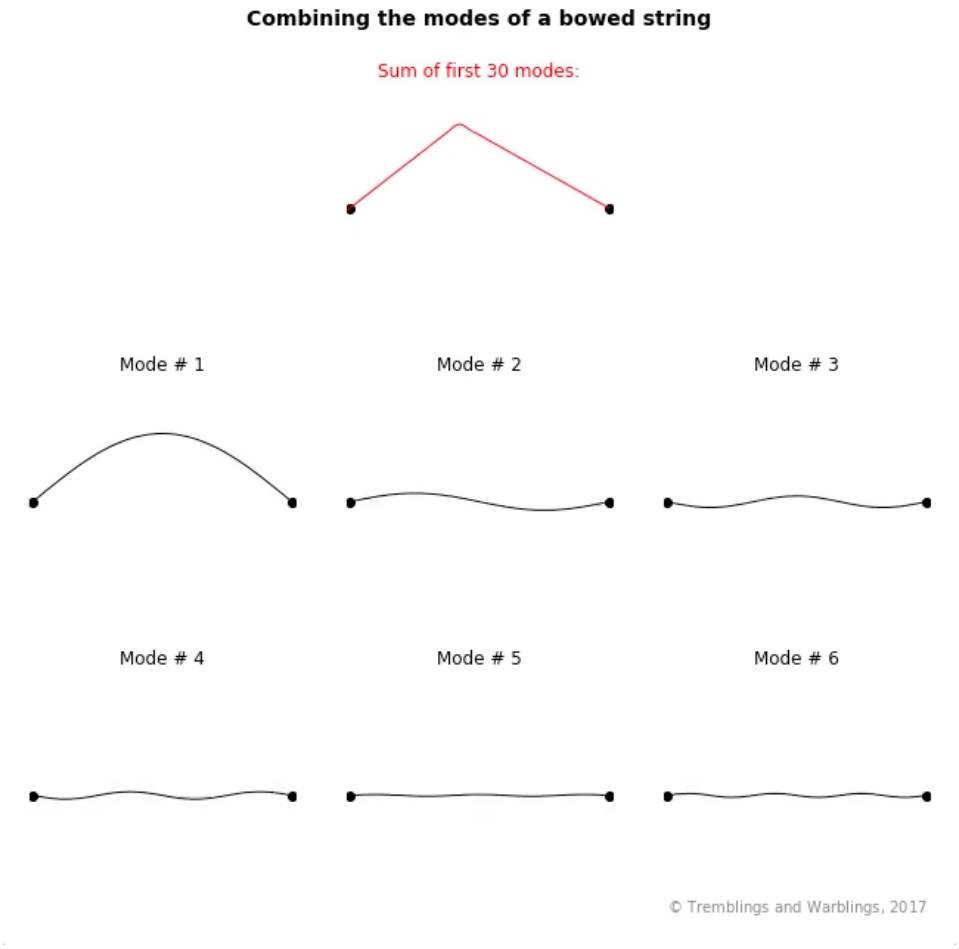


Figure 9.4: Still frame from an animation of a vibrating violin string, showing how sinusoidal sine waves add together to form a travelling kink.

Even though animations are more time-consuming to produce than still pictures, they often have better illustrative power. This particular animation emphasises the difference between the standing nature of the sinusoidal component waves, and the travelling nature of the resulting sum – something that is not evident from a still frame. It also makes it obvious that each component wave has a different frequency.

Since some of these demos and animations have not before been made by anyone else (as far as I am aware), they add a unique value to the blog.

To make the sound demos, I used discrete signal processing techniques such as Fourier analysis and harmonic detection. I wrote the code for the demos and animations in Python, and also made use of the `sms-tools` sound analysis and manipulation package [67]. My code is publicly available at `github.com/shtan/Blog`.

Using Media From Other Sources

Of course, I also use relevant media (images, sound samples and animations) which was created by others. It is important to only include media which I have permission to reproduce. Educational materials are often licensed under the Creative Commons licenses, and so can be reproduced under certain conditions. If no license is specified, I may directly contact the creator of the work to ask for permission to use it on my blog. I always include a caption with a citation of the source, containing the name of the work, its author's name, and the license name. The caption also includes links to the original source of the work, the author's home page, and the license text.

When necessary, and when the license permits it, I sometimes modify a piece of media. Figure 9.5 shows such a modified gif animation, where I have added some moving text and changed the axis labels to better explain my point. Note the caption describing the modifications (required under the license terms).

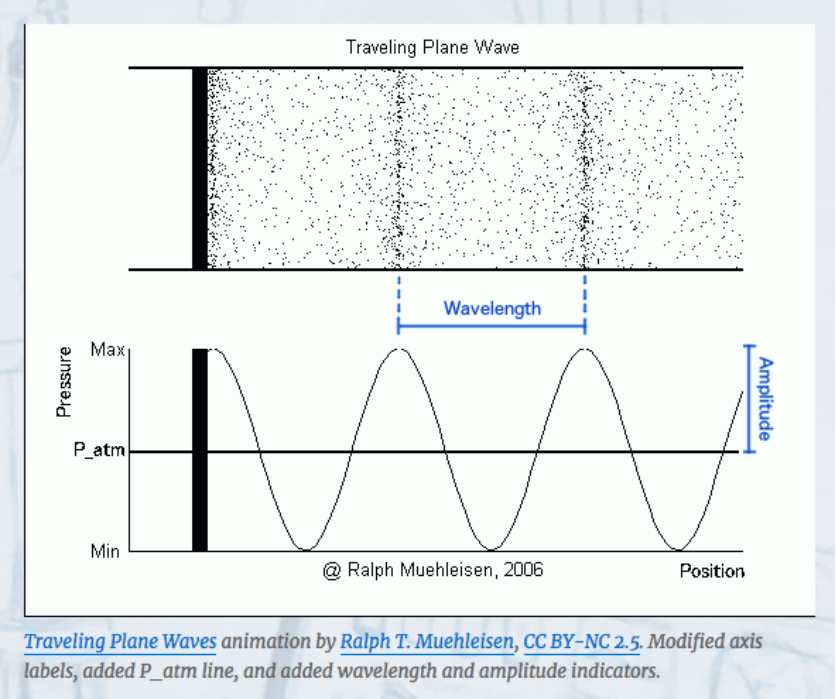


Figure 9.5: An animation which I modified. Its associated caption contains citations, links to the original source and license, and a description of the modifications made. Original animation from [68].

9.3.2 Post Length and Frequency

My blog posts tend to be long (1500 words on average) and well-crafted – I put quite a bit of thought into how to structure a post, and typically go through three or four drafts before publishing the final version. This contrasts with what is perhaps the conventional idea of blogging, where a gifted writer quickly pushes out posts about trending topics. There is nothing wrong with producing posts quickly, but this mode of writing is suited to neither my purpose nor my personality.

Making the media (recording sounds and coding up demos and animations) can also be very time-consuming. This is especially the case if the demos require a detailed understanding or a lot of trial-and-error to make, such as the animation of the travelling

kink in a string.

With the combination of researching a topic, crafting a well-written piece, and making demos, I have been writing about one post a month on average.

9.3.3 A Standalone Resource

One of my earliest goals for the blog was that it should be able to act as a standalone resource for someone who wishes to learn about the science of music and speech, without any background knowledge. It would have some of the characteristics of a book, where later posts build on background covered in earlier posts.

Therefore, my intention for the first few posts was to explain commonly-used concepts in the science of sound and music. These background posts are:

1. What is a Sound?
2. What Makes a Musical Sound Part 1 – The Basics
3. What Makes a Musical Sound Part 2 – Tone Quality and Spectra
4. The Fourier Transform and the Spectrum
5. What Makes a Musical Sound Part 3 – Envelope and Playing Techniques

These posts introduce concepts like frequency, harmonics, and spectra, which I use in almost every post that follows. Whenever a concept makes an appearance in a later post, I give a very brief description of it, and include a link back to the earlier post that explains it in more detail.

Once these concepts had been covered, I had more freedom to write about whatever I happened to be interested in at the moment. To help new readers to the blog find a place to start reading, I provide a list of these background posts in the blog's *About* section.

This goal of being a standalone resource, where later posts build on the earlier ones, means that posts are not necessarily timely or relevant to current affairs.

9.3.4 Depth and Level of Detail

To cater to readers who have different stomachs for detail, I implemented an "expandable section" feature using a WordPress plugin. Any text and media placed in an expandable section is collapsed by default when the page loads, showing only a descriptive heading. If the reader clicks on the heading, the section appears.

I use these sections for going into a topic in more depth. In this way, readers who wish to know more can read more, while other readers are not overwhelmed with detail. In particular, I avoid mathematical equations in the main text, relegating them to these expandable sections. (That said, I do not use many equations in the first place, since plots and animations can usually illustrate concepts better.)

The expandable section feature is also useful for going on "side detours" that are interesting, without breaking the narrative flow of the story. They have saved me several times when I had information that I just couldn't bear to leave out of a post, and which I would otherwise have had to force myself to give up.

9.3.5 Site Design

Aesthetic Aspects

I drew the images for the site header and background. Both pictures were inspired by Francis Bacon's *Sound Houses* passage, and feature somewhat imaginative sound-related contraptions which combine a steampunk and vintage style. Figure 9.6 shows the header image, which doubles as the blog's theme image, and automatically appears as the featured image when some external sites such as Facebook link to the blog.

Figure 9.6: Theme image of blog.

Navigation

The home page of the blog shows extracts of the most recent posts. I thought this would be preferable to the other common alternative of showing full posts on the home

page, since it allows readers to quickly scroll through and pick something that they find interesting. The site menu in the header bar also lets readers quickly navigate through the blog – it features links to the home page and *About* page, and includes a drop-down list of post categories. The right-hand bar has links to recent posts and to the blog archives.

I took special care to make sure that the mobile version of the site (for phones and tablets) displays just as well as the desktop version. In particular, I created a different site menu style for the mobile version, and adjusted the layout of the titles and text to make sure that nothing overlapped or was cut off. Having a good mobile site is important, because a large fraction of internet users use their phones to surf the web.

I assign each blog post a category, as well as a list of keyword tags. Both of these allow readers to easily find similar posts.

9.3.6 Use of Narrative

Many explanatory blogs use a direct style, delving straight into an explanation of a concept or subject. Some of my posts follow this mode, in particular the background-providing posts mentioned in section 9.3.3. For some of my other posts, however, I decided to experiment with incorporating some storytelling into the piece.

This was particularly appropriate for articles about some historical piece of technology. In the post *The Mechanical Talking Head: An Early Speech Synthesiser*, I described a mechanical speech synthesiser called the *Euphonia* and how it worked. In the process, I also explained the physics behind the human voice. I wrapped all of this within the framework of a story about the *Euphonia*'s inventor, Joseph Faber, giving details about

his life, work, disappointments, and legacy.

As described in Section 6.2, telling stories that show the human side of scientists makes the audience care more about their work. This type of narrative also provides an easy way to draw in the audience in the first paragraph of an article – by vividly describing a scene. As an example, here is the first paragraph of the talking head post:

In a dimly-lit room in the back of London's Egyptian Hall, a few curious people had gathered. Each had paid a shilling for the privilege of seeing the object standing in the centre of the room. A grotesque device it was – the mask of a woman's face, framed in the fashionable ringlets of the day, mounted on a frame which was attached to a piano-like instrument. Behind the keys sat a doleful-looking man whose clothes, though fine, had seen better days. The man placed his fingers on the keyboard and moved his foot on the pedal. The automaton's mechanical lips parted, and a spectral voice issued thence: "`Good morning, ladies and gentlemen.`*"*

Another way to add interest to an explanation of a scientific concept is to focus on some fascinating application that it has. For example, my post *Why Inhaling Helium Makes You Sound Like a Chipmunk* uses the well-known phenomenon of "helium voice" to explain the importance of the formants in our vocal spectra. For such posts, the phenomenon itself can be used as an attention-grabbing hook in the first sentence. In the helium article, I started the post by immediately presenting a video of the King's College Choir using a helium balloon to hit high notes (as part of an April Fool's Day joke).

9.4 Analysis of One Blog Post

All of this stuff that you've been telling me is really interesting. I know you already gave some examples of what you were talking about, but I think it would be really helpful for me to see these things in action in a post. Could we maybe go through one of your posts so that I can see how you apply these techniques?

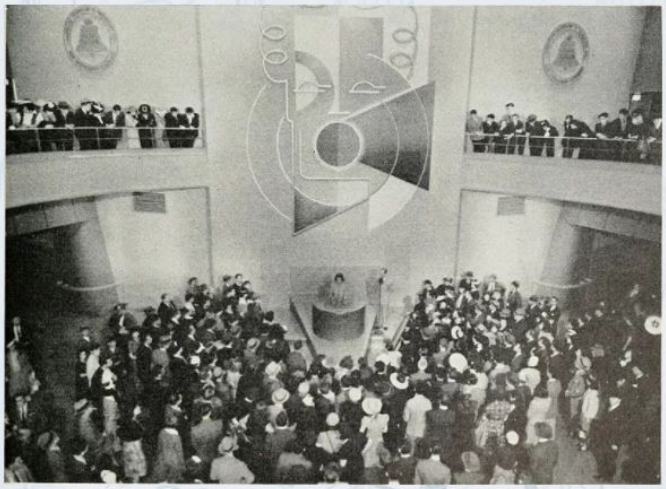
Sure, let me pull up a post on my laptop... Here, let's look through this one. I first wrote a version of it for a science communication workshop called ComSciCon last year, and had it critiqued by two fellow students and one professional science writer. I've since improved it based on the feedback.

This is a post about the first electronic speech synthesiser and how it worked. It is unique because I created a digital reconstruction of the synthesiser for the post, by generating source sounds and applying digital filters.

Tremblings and Warblings

MENU

Meet Siri's Great-Grandfather, the Voder



The Voder being demonstrated at the 1939 New York World's Fair. From the Bell Telephone Magazine, no known copyright restrictions.

Mrs Helen Harper, of Jamaica, New York, had an unusual ability — she could play the human voice with her fingers.

It was 1939, and the world's first electronic speech synthesiser had just been invented at Bell Laboratories in New Jersey. The *Voder*, as it was named, consisted of a keyboard connected via a bunch of circuitry to a loudspeaker. Under the nimble fingers of Mrs Harper, the *Voder* spoke and sang to audiences at performances around the country.

Attention-grabbing title. Mention Siri, whom most readers are familiar with.

Eye-catching image at the beginning of the post. It doubles as a feature image for the post, which accompanies it on the blog home page, as well as on some external sites (like Facebook or blog aggregators) which link to this post.

This first sentence piques the reader's curiosity, and also emphasises the human element.

An introductory paragraph, expanding on the story. It also contains many keywords for the post, such as "first electronic speech synthesiser" and "Voder". Having keywords near the beginning of a page improves its ranking with search engines.

A contemporary video of the Voder in action. It has a somewhat metallic voice, but is easily understandable; and Mrs Harper is able to use expression to convey different nuances with the same words.

Two centuries earlier, inventors had already created mechanical speaking devices that modelled our mouth, vocal cords, and lungs. We saw one of these, the *Euphonia*, in a previous post. But the *Voder* was unique in being the first machine to speak by electronic means — the precursor to the modern computer-generated voices that are so ubiquitous today. In this post I'll explain how the *Voder* spoke, with the help of some demos which imitate its workings.

The Voder's Sound Source

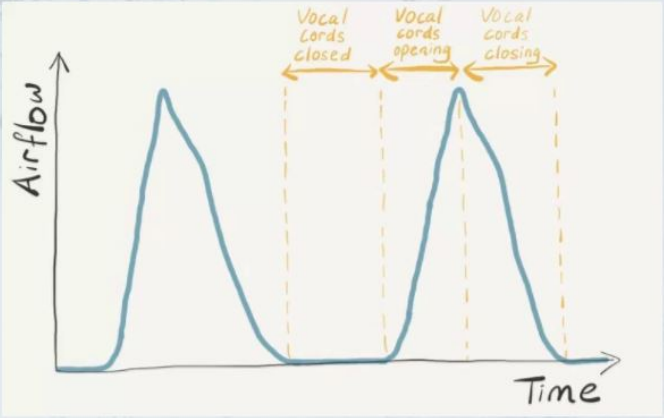
Just how could a bunch of electrical circuits produce the sound of human speech? To understand this, let's first briefly recap what we saw in a previous post about how the human voice works. When we speak, air flows from our lungs up through our throat and between our vocal cords. The vocal cords vibrate, opening and closing hundreds of times per second, and allowing the air to escape in pulses.

Short video for readers to see and hear the *Voder* in action. A caption which describes the video content, in case a reader doesn't feel like watching it.

One of my science communication professors calls this the "coat-hanger paragraph". It joins the introductory material to the body of the post, giving a quick summary of the post content. It also hints at why this information might be relevant to the reader, by mentioning the modern speech synthesiser technology that we are all familiar with.

Asking a pertinent question (and then answering it yourself later) serves to focus the reader's attention.

To mimic this source of sound electronically, we need to know how the air pulses are shaped. Here's a graph of what the airflow through our vocal cords looks like over two open-and-close cycles:



Graph of vocal cord airflow volume against time

This hand-drawn plot is an exception to my usual rule of only using plots of real data. In this case, I ran into technical problems while trying to produce a plot of the glottal airflow.

The *Voder* simulated this source of sound by generating an electrical wave of a similar shape. The wave sounds something like this:

1 Source sound imitating vocal cord buzz

00:00 00:01

Embedded audio file with a descriptive label.

This is what our vocal cords would sound like, too, if we heard them by themselves. But as we saw in an earlier post, this sound gets filtered by the vocal tract (our oral and nasal cavities). The vocal tract picks out certain frequencies and suppresses others, forming mounds in the spectrum called formants. These formants are responsible in part for the differences in tone quality between male and female voices, and also between individual speakers.

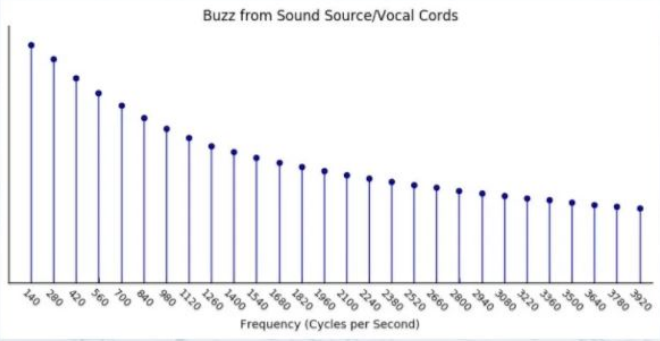
Links to previous posts which explain how sound is filtered by the vocal tract, and how this creates formants. In this paragraph, I also provided a brief explanation of these concepts, enough to allow a reader to understand the rest of the post without using the backlinks.

But formants play another important role — they allow us to distinguish between different speech sounds. By shaping our mouth differently, we can change its resonant frequencies and shift the formants around, thus producing different vowels.

The Voder's Electronic "Mouth"

But what about the *Voder*? It didn't have a squishy cavity made of muscle which it could shape to modify its sound — it was just a bunch of electronics! Instead, it created formants using electrical circuits called band-pass filters. Each filter picked out a certain band of frequencies, and suppressed others. There were 10 filters in total, each one connected to a key on the keyboard[1].

To better illustrate this, I've made a few demos that show how we can take the buzzing source sound we heard earlier, and change it into various vowels by picking out only certain bands of frequencies. Let's take a look at the spectrum of the buzzing sound to see what frequencies are present in it:



Headings break up the text into chunks. This helps the reader to structure their thoughts, and provides a kick of encouragement whenever they finish a section.

This is the actual spectrum of the buzzing sound presented on the previous page. It looks so idealised because this is a generated sound.

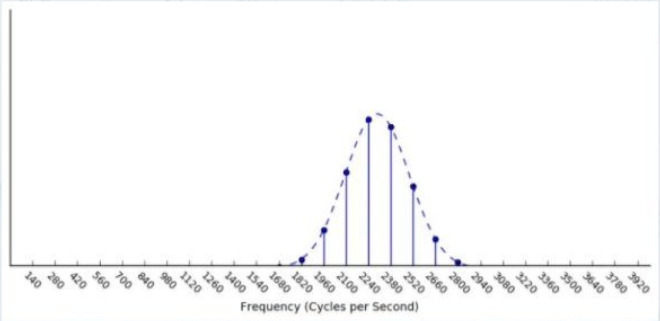
This particular example wave has a frequency of 140 Hz, which means that, had it been produced by vocal cords, the cords would have had to do their open-close cycle 140 times per second. From the spectrum, we can see that the sound also contains harmonics at 280 Hz, 420 Hz, 560 Hz and so on — multiples of the fundamental frequency of 140 Hz.

^ Click to Expand: A Sawtooth-Shaped Airflow

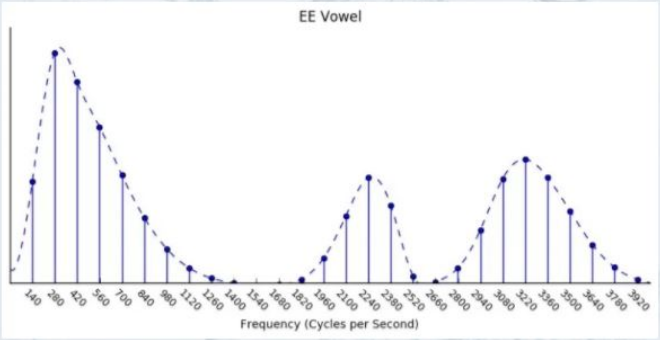
If you look at the graph of the airflow through our vocal cords that I showed earlier, you'll notice that it looks pretty triangular, instead of being smooth. This is quite important. As we saw in this article about the Fourier theorem, waveforms with sharp corners tend to have lots of strong high harmonics. We're going to apply band-pass filters that pick out bands of frequencies, corresponding to the resonant frequencies of the vocal tract, and some of these resonant frequencies are pretty high. In order to pick out these frequencies from the source sound, there has to be something there to pick out in the first place! If the source sound were nice and smooth with barely-existent high harmonics, there wouldn't be very much for the filters to act on.

An expandable section that provides additional information, without breaking the flow of the main text. The reader clicks on the heading to make the text in blue appear.

What happens if we take our buzzing sound, and apply a band-pass filter that picks out frequencies near, say, 2300 Hz? We get a spectrum that looks like this:



It turns out that, to make vowels, we need to apply two or three band-pass filters at once. Let's apply filters at 270, 2300 and 3200 Hz:



Here, I've been a bit fancy with the filters, and adjusted each of their heights, widths and shapes separately. (The *Voder* probably didn't have this level of control.) What does this filtered sound sound like?

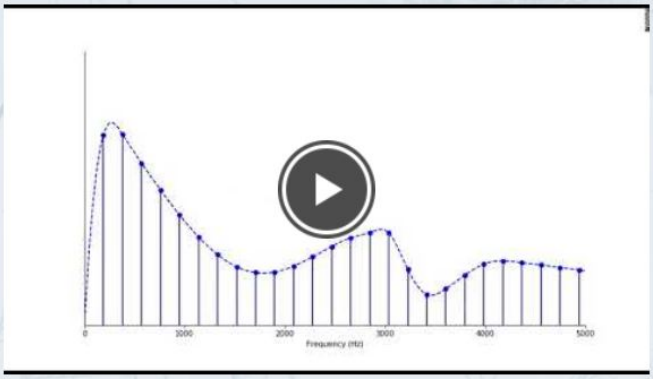
1 Filtered EE vowel

00:00 00:01

Starting from simple examples and building up to a more complicated demo allows the reader to follow along more easily.

Interspersing text and media prevents the dreaded "wall-of-text" effect that makes someone stop reading.

That sounds like an "ee". A bit metallic, perhaps, but it certainly sounds more like an "ee" than the original buzzing sound. If we move the three filters around, we can get different vowel sounds. You can hear this in the following video (it has a different fundamental frequency, but the concept is the same):



I don't know about you, but I find that pretty exciting.

Of course, vowels aren't the whole story — to pronounce words, we need to make consonants, as well. We produce these by constricting some part of the vocal tract to restrict the air flow. To make different consonants, we can change the place where the constriction occurs, whether or not we vibrate our vocal cords, or the length of the sound.

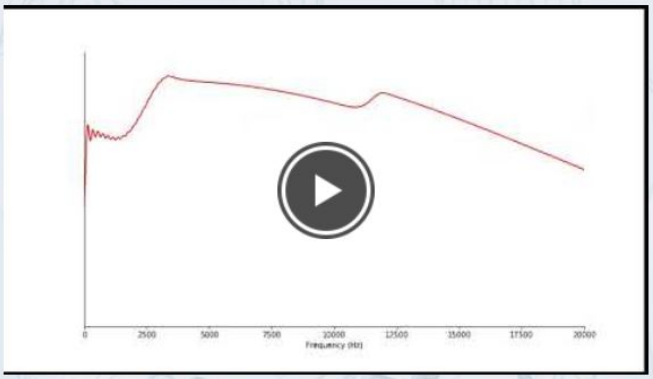
Let's take a closer look at these three factors, starting with the last one. Sounds like "s" and "zh" can be arbitrarily long, while sounds like "t" and "b" are short. The latter are called **stops**, because we produce them by actually stopping the airflow for a fraction of a second, and then releasing a burst of air. We can simulate a stop using a burst of high-frequency sound, and the *Voder* had special keys to do so.

Writing in first person, and stating an opinion, adds a touch of personality to the post.

I originally wrote the above section about vowels as an explanation of how the *human voice* produces vowels. However, both students who read this post at ComSciCon thought that this long foray into human vowel production was a distraction from the story of the *Voder*. I found this surprising, since in my mind, this was the main part of the explanation of how the *Voder* worked! I realised that I could re-work this section into an explanation of how the *Voder* itself worked, all while using the same media and explaining the same concepts.

Consonants can also be **voiced** or **unvoiced**. In voiced sounds, the vocal cords vibrate, while in unvoiced sounds, we hold the vocal cords open and let air pass through unhindered. (If you put your finger on your throat while pronouncing the sounds "z" and "s", you'll feel the difference.) The *Voder's* wrist bar allowed the operator to switch between a buzzing source to produce voiced sounds, and a white-noise source to produce unvoiced sounds, as you heard in the video at the beginning of this post.

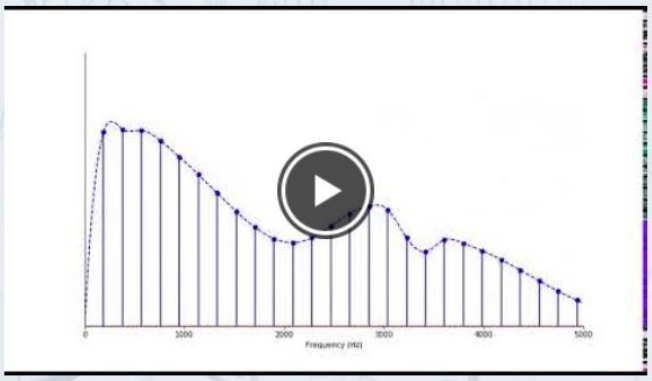
What about where we constrict the vocal tract? Compare the sounds "s" and "sh" — in the former, the air passes between the top of the tongue and the roof of the mouth. In the latter, you curl your tongue up a bit, and let the air pass under it. Just like with vowels, the different mouth shapes give the two sounds different formant frequencies, as you can see in the following video.



I avoid the use of jargon (though it can be tricky to know what words most people know or don't know!). Using unfamiliar words is sometimes inevitable, however. In these cases, I print the word in bold type so that readers know that it's supposed to be unfamiliar, and don't feel bad for not understanding it. The bold font also marks the word out as important vocabulary, and makes the word easier to find when a future post back-links to this concept.

Youtube videos are easy to embed in WordPress blog posts. I use this method even for videos that I made myself, because it allows me to host the video on YouTube instead of on the blog itself. This reduces the size of the blog and improves loading speed.

Now let's put the consonants and vowels together to say "she saw me":



That sounds rather flat and robotic... To make the voice more expressive, we can continuously change the pitch of the voice by changing the fundamental frequency of the voiced buzzing source. (The *Voder's* operator did this using a foot pedal.) By putting the stress on different words, we can change the emphasis of the sentence:

If a picture speaks a thousand words, an animation with 300 frames might speak 300,000. These ones have both sound and video, and show how changing the formants in the spectrum causes different speech sounds. These animations were probably the most challenging ones I've made so far, partly because it took a lot of trial and error to get the synthesised words to sound right. I analysed real vocal recordings to measure the positions of the formants for different vowels and consonants, and reproduced these spectral shapes using multiple asymmetric Gaussian filters. The speech sounds then had to be combined together, with the right duration and volume for each sound, proper cross-fading between sounds, and realistic varying of pitch to create expression. The result doesn't sound as good as the original *Voder*, but the animations made it much easier to explain how it worked.

This demo I made doesn't sound as good as the original *Voder* — I think it sounds creepier. It also has particular difficulty with the "m" in "me", and the expression is perhaps somewhat exaggerated. Evidently, the *Voder's* formant filters were very well-tuned — its inventor, Homer Dudley, had previously spent time analysing the human voice using the Vocoder, another machine that he had developed.

Again expressing an opinion.

Learning to Play the Human Voice

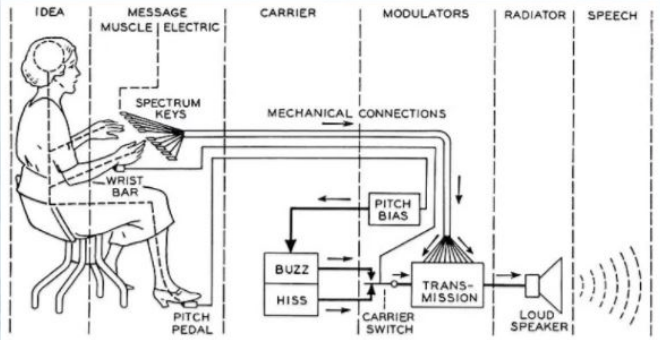


Fig. 8—Schematic circuit of the voder.

Schematic of the Voder. From the Bell System Technical Journal, no known copyright restrictions.

Picture from an external source, with a caption containing links to its source and information about licensing restrictions.

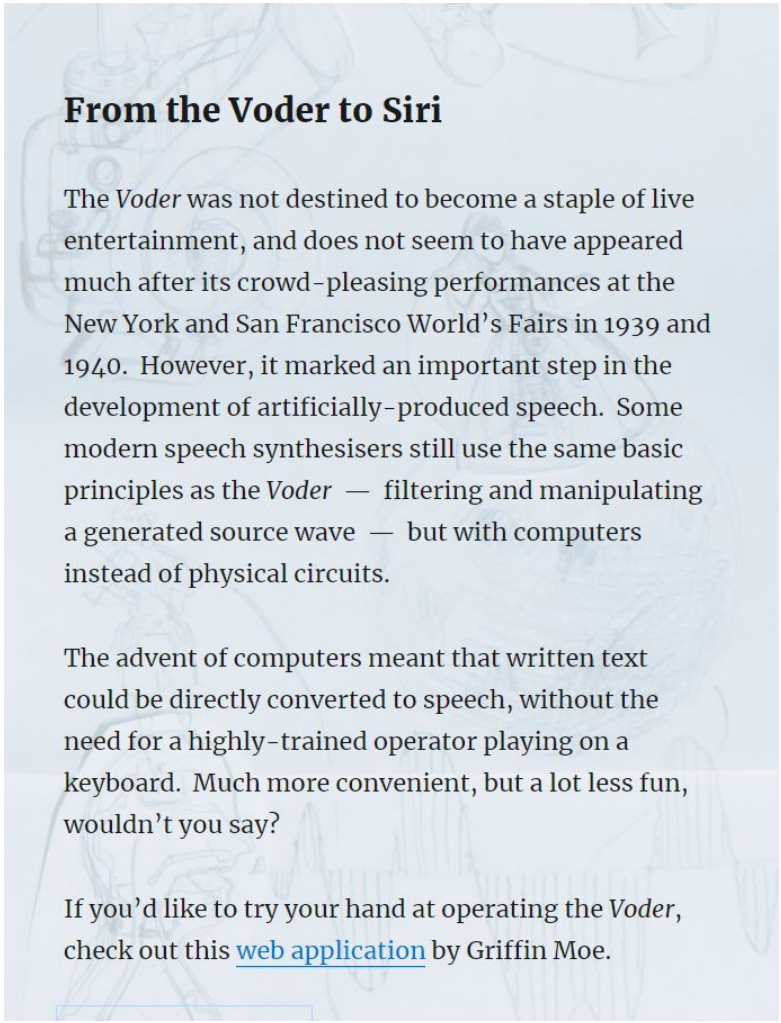
Not only did the *Voder's* electronics have to be tuned to perfection — it had to be played to perfection, as well. The operator had to memorise the formant frequencies for each sound, and press the correct combination of keys, pushing each one down just the correct amount — all at the rate of normal speech! She also had to produce the stop sounds, and deal with the wrist bar and foot pedal. As Mrs Harper explained:

"In producing the word 'concentration' on the Voder, I have to form thirteen different sounds in succession, make five up-and-down movements of the wrist bar, and vary the position of the foot pedal from three to five times according to what expression I want the Voder to give the word. And of course, all this must be done with exactly correct timing."

When Stanley Watkins, who programmed the *Voder*, first sat down at the instrument, he attempted to make it say "yes". Everyone else present thought it had said "peanuts". To find someone who would be able to learn to play it well, he searched among the Bell Telephone Company's telephone "girls", finally selecting Mrs Harper based on her finger agility, ability to distinguish sounds, and appearance(!)[2]. She later helped to train a number of other women to operate the *Voder*; of the 300 women that the Bell Telephone Company tried out, only 28 got good at it[3]. It took each of them about a year of constant practice to be able to produce intelligible, natural-sounding speech on the machine.

Direct quotes add liveliness and a human element to the post.

It is not that common to cite references in blog posts or websites, but I do so both for completeness, and to provide my readers with links to more information.



From the Voder to Siri

The *Voder* was not destined to become a staple of live entertainment, and does not seem to have appeared much after its crowd-pleasing performances at the New York and San Francisco World's Fairs in 1939 and 1940. However, it marked an important step in the development of artificially-produced speech. Some modern speech synthesisers still use the same basic principles as the *Voder* — filtering and manipulating a generated source wave — but with computers instead of physical circuits.

The advent of computers meant that written text could be directly converted to speech, without the need for a highly-trained operator playing on a keyboard. Much more convenient, but a lot less fun, wouldn't you say?

If you'd like to try your hand at operating the *Voder*, check out this web application by Griffin Moe.

If I had been writing a standalone article for a magazine, I would have expanded more on this ending, making it more relevant to the reader's daily experiences of modern speech synthesisers. I didn't do this here, because I'd already written several other posts about speech synthesis before this one, so it is already a theme of the blog.

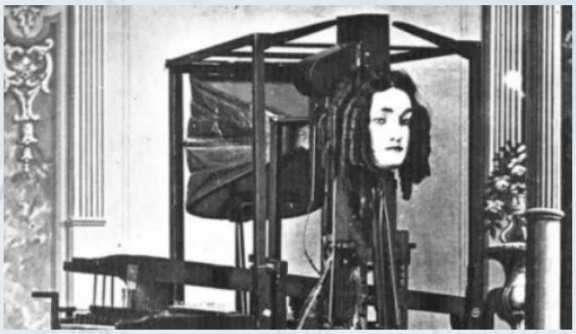
By linking to interesting external sites, I introduce my readers to good resources, and also help to build a network between similar sites.

References

1. Talking Heads: Simulacra. Haskins Laboratories: The Science of the Spoken and Written Word. http://www.haskins.yale.edu/featured/heads/SIMULACRA/voder.html.
2. Otto D. Helen Harper Playing the Voder will be the Hit of the Fair. *Long Island Star-Journal*. http://fultonhistory.com/Newspaper%2014/Long%20Island%20City%20NY%20Star%20Journal/Long%20Island%20City%20NY%20Star%20Journal%201939/Long%20Island%20City%20NY%20Star%20Journal%201939%20-%201521.pdf. Published January 10, 1939.
3. The Voder – Homer Dudley (Bell Labs) 1939. YouTube. https://www.youtube.com/watch?v=5hyI_dM5cGo. [Source]

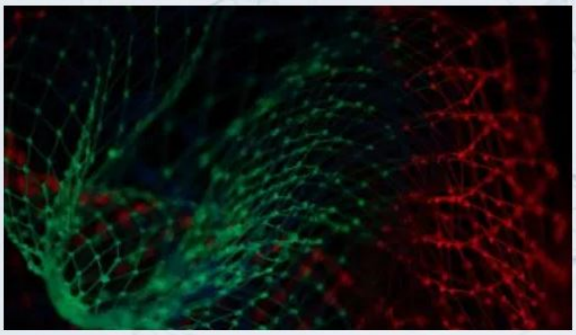
Share and Like buttons encourage publicity.

Related



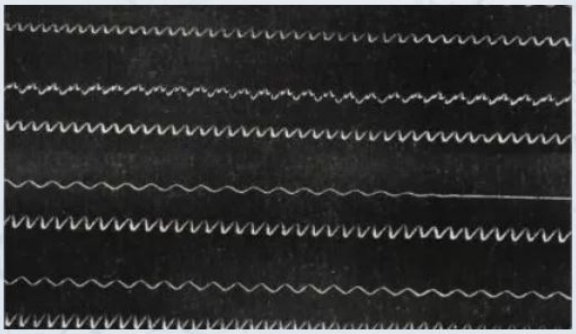
The Mechanical Talking Head

June 6, 2017
In "Historical"



WaveNet: Micro-Managing the Way to Realistic Synthesised Speech

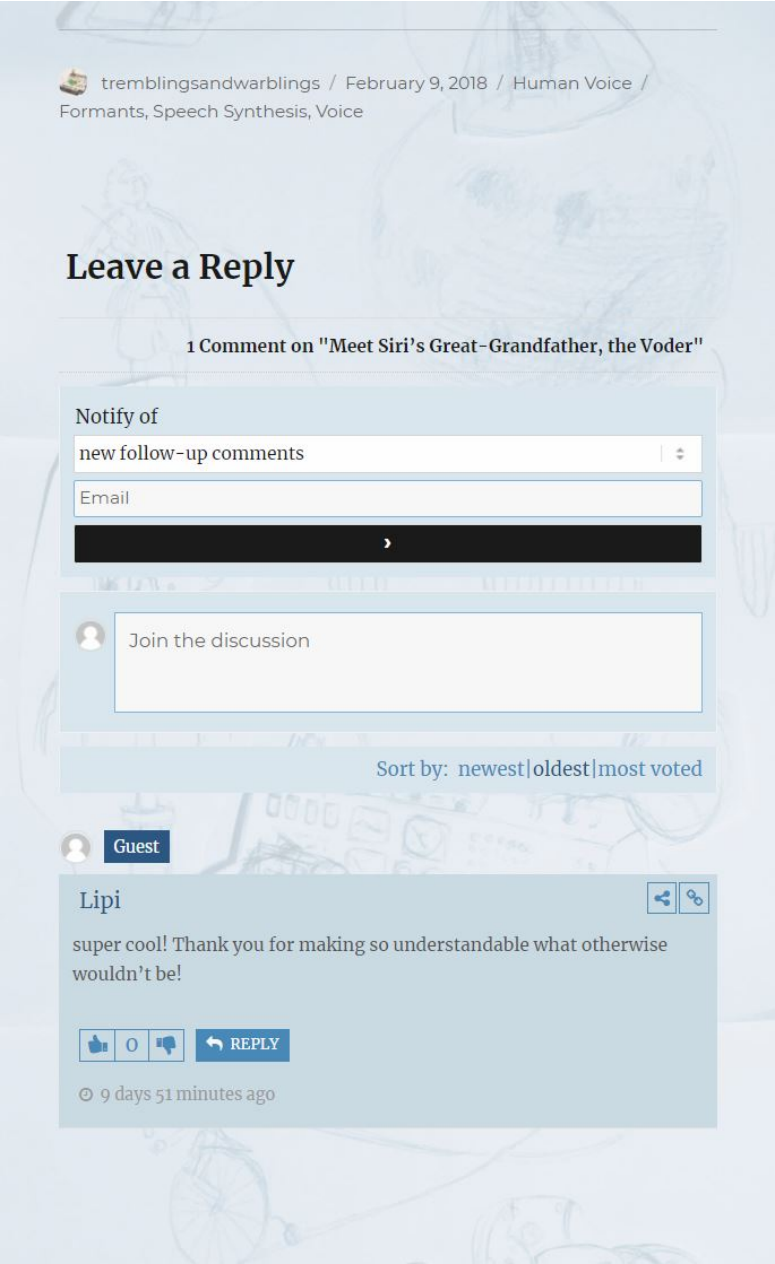
November 6, 2017
In "Human Voice"



The Voice in the Soot: Humanity's Earliest Known Recording

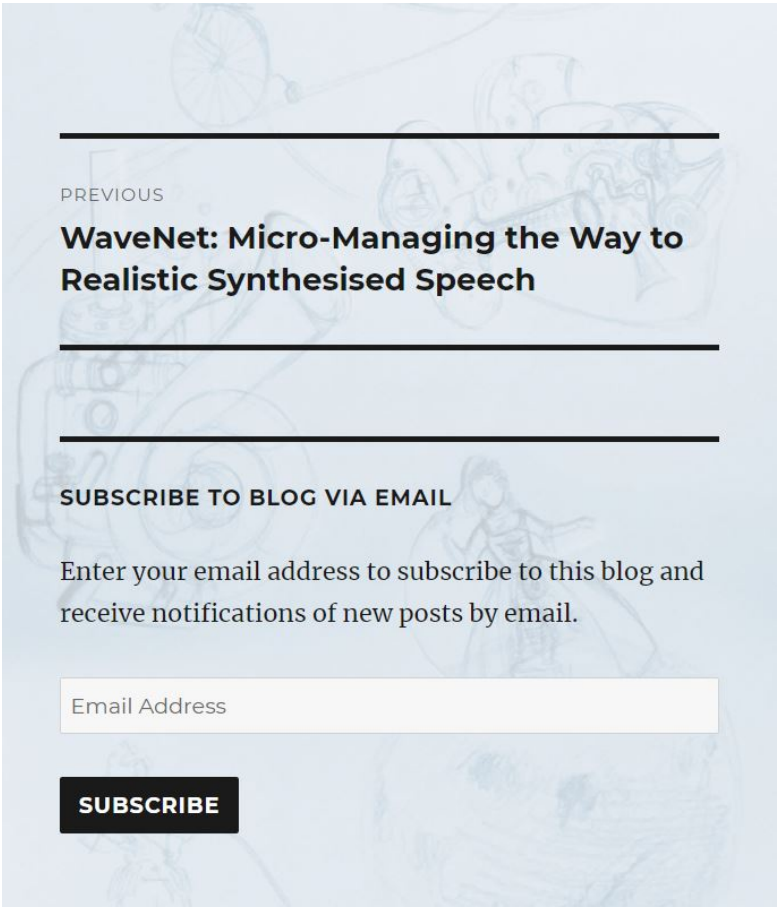
January 24, 2017
In "Historical"

Present related posts to help readers decide what to read next. Notice that the feature image for each post is shown, to pique the reader's interest.



Post category and keyword tags

The comment section encourages discussion and could eventually be useful for building a community of readers. I provide commenters the option to enter their email in order to receive notifications of replies.



Links to previous and next post

The "subscribe" box appears both at the bottom of a post, and in the sidebar (which I haven't shown here)

General Comments on Narrative Elements

In this post, I used Mrs Helen Harper and her ability to play the *Voder* to add a human element to the story. It is not a particularly pervasive narrative thread – she is introduced in the beginning, and appears again near the end. James Dacey, the professional science writer who gave me feedback on this piece, suggested that I should work more of Mrs Harper's story into the post – perhaps using it to build up dramatic tension throughout the article, making the reader wonder whether she would ever be able to learn how to play the *Voder* properly.

I had trouble implementing this suggestion, in part because I could not find very much information on Mrs Harper's story. Contemporary accounts of her experience with the *Voder* were written in the Gee-Whizz style of the 1930's and 1940's (see section 3.4) – they were more focused on promoting the amazing technology than on describing the difficulties faced in developing it. In addition, focusing too much on this story would have made it difficult to fit in the detailed explanations of how the *Voder* worked, as well as the *Voder* reconstruction animations, which were the main point of the post.

In the end, how much to emphasise the human story depends on the purpose of the work. At a workshop I attended, physicist and science writer Tasneem Zehra Husain expressed the opinion that science writing just needs to have small mentions of human stories here and there – it doesn't have to be *dripping* with emotion throughout the piece. I thought that this seemed to be an appropriate level of emotiveness for a blog piece with an explanatory goal. Had I been writing a feature article for a more generic newspaper or magazine, I would probably have built more on the drama of the story.

9.5 Building an Audience

`It sounds like you're really having fun making these posts!`

I am! And I'm learning a lot of stuff on the way too. The one thing that I don't find so enjoyable is having to promote the blog!

`Haha, is it very tedious?`

It's tedious in a different way from making the posts. But despite that, I do make an effort to build an audience.

`I noticed in your example post that you have a comments section for each post.`

Yeah, that helps to build a community, and if a reader has any questions about a post they can ask me directly about it.

`You also have an option to allow people to subscribe to your blog. Does this mean that they get an email whenever you post a new article?`

Yep. I post only every month or so, so it's a good way to prevent people from forgetting about the blog. Besides, the blog name, *Tremblings and Warblings*, is somewhat difficult to remember, since it contains rare words. This is generally not a good idea for website names, but I went with it because I couldn't think of a better name, and because I liked the sound of it. I also use some more advanced

techniques to promote the blog, as I'll describe in this section.

9.5.1 Search Engine Optimisation

Search Engine Optimisation, or SEO, is the process of making sure that someone can find your website using a search engine such as Google. This usually involves applying techniques to make the site rank higher on relevant searches.

I used the Yoast SEO WordPress plugin for many of my SEO tweaks. I started by allowing my site to be indexable by search engines, and by creating an XML sitemap, which makes it easier for search engines to crawl through all the pages and links in the site.

I also optimise each post individually by doing the following:

- Writing a compelling meta-description (the little blurb that appears below the post title in search engine results). See Figure 9.7 for an example.
- Removing stop words from the post URL. Stop words are words like "the" and "of" which have little content.
- Providing "alt text" for images. Alt text is text that is associated with each image, containing a brief description of the picture. It appears on the page if the image fails to load, and is also useful for visually-impaired readers who use automatic voiced screen readers. In addition, search engines use alt text when crawling the website and when indexing images, so it's a good idea to include keywords in the alt text.

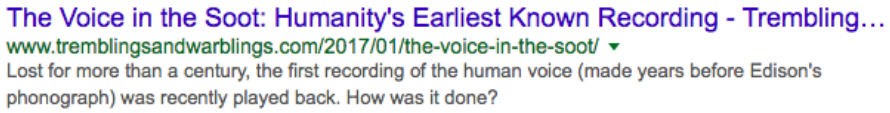


Figure 9.7: Meta-description for a post, as it would show up in a search engine search.

SEO vs. Good Writing

Sometimes, I feel that following the rules of optimising for search engines would get in the way of good writing. In such cases, I always prioritise the writing. This is particularly the case with SEO's obsession with keywords. For example, one SEO rule is to include keywords close to the beginning of the post, in the first paragraph. This is sometimes not feasible when one is using the first paragraph to set a scene. For example, the post about the mechanical talking head does not contain the obvious keywords "Euphonia" and "speech synthesis" in the first paragraph; and similarly for the *Voder* post. Forcing the keywords into the first paragraph would have rather spoiled the scene.

It is also difficult to optimise the blog for more general keywords that someone may use to search for it. For instance, the search query "physics of music" is very relevant to this blog – but none of the individual posts contains this phrase. They don't even contain the word "physics"! This is because the individual posts are about more specific topics, and peppering them with words like "physics" would add nothing to the content. I do, however, use this phrase in the list of categories, and in the *About* page.

9.5.2 Speeding Up the Site

Today's internet users have short attention spans, and tend not to sit around waiting for a slow site to load. Faster load speeds can also help to improve a site's page ranking

with search engines. I therefore spent quite a bit of effort on reducing the load time of the blog.

Optimising Media

Part of the challenge lies in the fact that the blog uses a lot of media such as images, sound and animations, which are large (in terms of file size) and thus load slowly. To mitigate this, I first optimise all the media before uploading it. This involves reducing the pixel size of images (including the background and header images). Since the content width of the main text area of my blog is only about 800 pixels, images do not need to be wider than this. In addition, I use image optimising software on images and gif animations – the software intelligently re-encodes the media file to reduce its size. I also optimise audio files by using low bitrates and a compressed file format.

For animations with sound, I usually upload them to YouTube and embed the YouTube video in the blog post. This means that the file is not hosted on my site, and so does not slow down the page load. The disadvantage of this procedure is that YouTube always displays videos using a certain aspect ratio. When an animation does not follow this aspect ratio, I have to directly upload it onto the blog, so that it displays with the correct size and with a good resolution.

In addition, I adjust the page settings so that the images and audio do not load immediately, but only after the text. I also wrote CSS code for the YouTube video embedding, so that only the thumbnail picture is displayed when the post first loads. The video itself only loads when the reader clicks on it. This reduces the initial load time of the page.

Caching and Content Delivery

I further speed up the site by using a caching plugin, W3 Total Cache. This allows a visitor's browser to store a copy of the site, so that it loads faster the next time they visit. I also use CloudFlare, a Content Delivery System that uses servers placed around the world to deliver content more quickly to visitors. Finally, I reduce the number of HTTP requests during a page load by minifying (combining and removing unnecessary code from) javascript and CSS files, if doing so does not affect the site's formatting.

Effect on Load Time

After implementing these changes, the load time of my home page went down from about 7 and 10 seconds (when accessed from California and Sweden respectively) to about 4 and 6 seconds.

9.5.3 Publicity

Promoting the blog through external sites serves the dual purpose of reaching a wider audience, and creating links to the blog from these sites which help to improve its search engine rankings. I publicise the blog through the following avenues:

Content Aggregators

I submitted the blog to ScienceSeeker, a content aggregator website that collects science blogs. It displays new blog posts to users, arranging them by blog category so that users

can read about topics that interest them.

YouTube Channel

Tremblings and Warblings has its own YouTube channel, where I upload the animations that I embed in blog posts. In the video description on YouTube, I provide a link back to the relevant post. This has been effective at drawing visitors to the blog.

Social Media

Each time I publish a new post, I share it on Facebook and LinkedIn, the social media sites that I use. I have considered expanding my usage to other sites such as Twitter and Reddit (which have a larger reach). However, the culture of these sites dictates that one must actively participate in discussions and post good content from many sources, before one can promote their own work. I ultimately decided that it would be too time-consuming to build up a reputation in this manner.

9.5.4 Visitor Stats

While I started publishing posts in January 2017, I only started publicising the blog in October 2017. Between then and February 2018, there were approximately 300 visitors to the blog, with about 680 page views between them. (These numbers are approximate, because I had to estimate my own visits to the blog, and subtract it from the stats aggregated on the website.) I also have 9 email subscribers.

The most common way that people found my blog was through search engines (83

referrals), Facebook (76 referrals) and YouTube (42 referrals).

CHAPTER 10

PERFORMANCE OF THE TOP RECONSTRUCTION FITTER

It's your turn to talk now... I think I've said enough about my blog! Tell me more about the top reconstruction fitter. What did you do with it after you got it to work and converge and all that?

```
Well, I tested it to see how it performs.
```

Tested it? On simulated Monte Carlo data?

```
Yep.  The test was done in two stages -- first, on clean events
   generated by Madgraph, that is, events that hadn't gone through the
   hadronisation and showering and detector simulation steps.  Then,
   I applied the fitter to more realistic events, which had been
   produced by the full MC simulation, and went through the process of
   calculating discriminator distributions and determining the limits
   on the ttbarH process, with and without the reconstruction.
```

10.1 Single-Leptonic $t\bar{t}H$ Events from Madgraph

The top reconstruction fitter was first applied to 20000 MADGRAPH-generated single-leptonic $t\bar{t}H$, $H \to b\bar{b}$ events, where one W decays into an up and a down quark and the other into an electron and a neutrino. The $t\bar{t}H$ system and its daughters were the only particles in each event.

To simulate the effect of measurement uncertainties and detector resolution, the generated events were randomly smeared according to a Gaussian probability distribution function. Jet p_T resolutions were set to $\sqrt{p_T}$, while jet angular resolutions (for η and ϕ) were fixed to 0.01. Lepton p_T resolutions were set to $0.01\,p_T$ and lepton angular resolutions to 0.001. The masses of the particles remain the same. These smeared values were fed into the top reconstructor, with the correct permutation of b quarks specified. The resolutions used to smear each quantity were also fed into the fitter, as the uncertainty σ.

The end-state daughter particles in a single-leptonic $t\bar{t}H$, $H \to b\bar{b}$ event are two b quarks from the tops, one lepton, one neutrino, two light quarks from the hadronically-decaying W, and two b quarks from the Higgs. Each of these particles is defined by four kinematic quantities: p_T, η, ϕ and mass (or alternatively, p_x, p_y, p_z and energy).

To examine the result of the kinematic reconstruction, we would like to compare three datasets – the MADGRAPH-generated particles, the smeared particles, and the particles output by the kinematic fitter. For the MADGRAPH-generated dataset, all four of the defining quantities are known for all of the particles, including the parent W's, tops and Higgs. The smeared dataset, on the other hand, simulates measured data from the detector. For this case, the four defining quantities for all end-state daughter particles except the neutrino are known. We can take the x- and y-components of the neutrino momentum to be equal to the MET (since there is only one neutrino in the single-leptonic case). The z-component of neutrino momentum is unknown, and is taken to be zero. We can then calculate the W by summing the momenta of the lepton and neutrino, and the top by summing the momenta of its daughter b and W.

For the best-fit case, the fitter has estimated new values of p_T, η and ϕ for all the

end-state daughter particles, including the neutrino. We take the mass of each of these particles to be equal to their mass in the original smeared dataset (this is also equal to the generated value of the mass). Using these estimated momenta for the daughters, we can also calculate the parent W's, tops and Higgs.

10.1.1 Residual Plots

As hinted at in section 8.2.3, a good measure of the performance of the top reconstruction fitter would be to compare the quantities

$$p_{\text{fit\_gen}} = p_{\text{fit}} - p_{\text{gen}} \tag{10.1}$$

and

$$p_{\text{smeared\_gen}} = p_{\text{smeared}} - p_{\text{gen}} \,. \tag{10.2}$$

Here, p is some kinematic quantity, such as momentum, mass or energy. The subscript *gen* refers to the generated (ground-truth) value, *smeared* is the value we feed into the fitter, and *fit* is the best-fit value returned by the fitter.

Figures 10.1 through 10.3 show histograms of $p_{\text{fit\_gen}}$ and $p_{\text{smeared\_gen}}$ for various kinematic variables for different particles in the $\mathrm{t\bar{t}H}$ system. Only events for which the fit converged (about 91% of events) are shown.

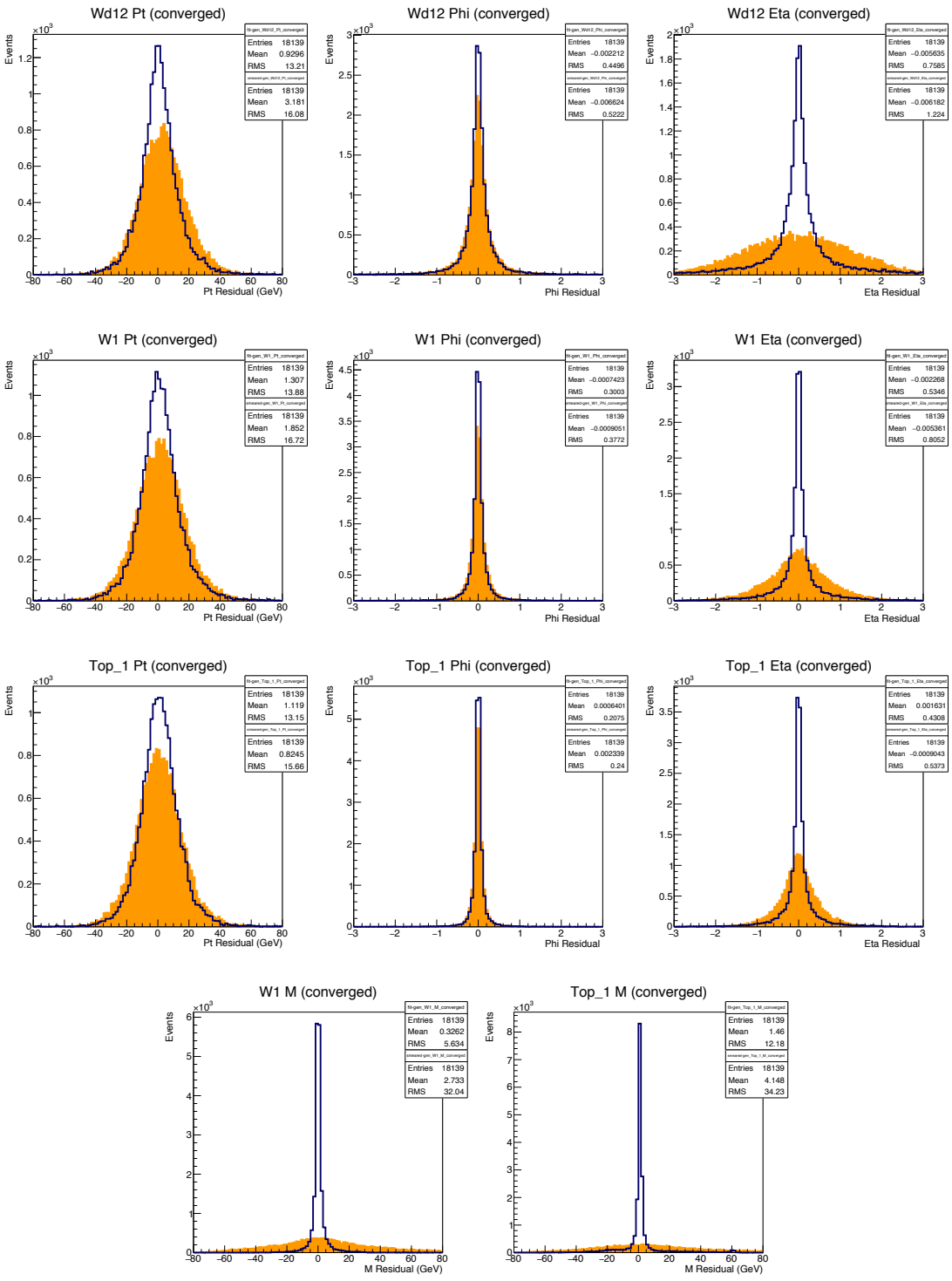


Figure 10.1: p_T, ϕ and η values for the neutrino and leptonically-decaying W and top, as well as W and top mass, for events for which the fit converged. Blue: $p_{\text{fit\_gen}}$; Orange: $p_{\text{smeared\_gen}}$. The neutrino is labelled "Wd12", and the leptonically-decaying W and top are labelled "W1" and "Top_1" respectively.

Figure 10.1 involves the particles that are directly dependent on the neutrino: the neutrino itself, and its parent W and top. We can see in the plots that the top reconstruction fitter yielded a dramatic improvement in the momentum estimates of the neutrino and leptonically-decaying W and top – the distributions of $p_{\text{fit\_gen}}$ are much narrower than those of $p_{\text{smeared\_gen}}$, indicating that the former has a significantly lower absolute value on average. This effect is particularly marked for the η values of the particles, as well as the W and top masses. This is unsurprising in both cases – the η values depend directly on the z-component of the neutrino momentum, which was unknown before the fit; and the W and top masses were part of the constraints used in the kinematic fitting method.

Figure 10.2 shows the results for the top-system particles whose momenta are not directly dependent on the neutrino. For the end-state particles among these (the two b quarks and two light quarks from the hadronically-decaying W) I have only shown the p_T variable, since the mass is untouched by the smearing process, and ϕ and η have such small resolutions as to be almost unchanged by the smearing or fitting process. For the hadronically-decaying top and W, I have included both the p_T and mass plots (the angular variables are again mostly untouched by the smearing and fitting processes). I have not included the lepton, because again its p_T and angular resolutions are too small for its momentum to have changed much.

Once again, we see that the fit has dramatically improved our estimates of these values, even though they are not directly affected by our estimate of the neutrino momentum.

The non-top system, however, reacts differently to the fit. Figure 10.3 shows plots for the Higgs and its two daughter b quarks. We can see that the top reconstruction process does not affect these distributions very much.

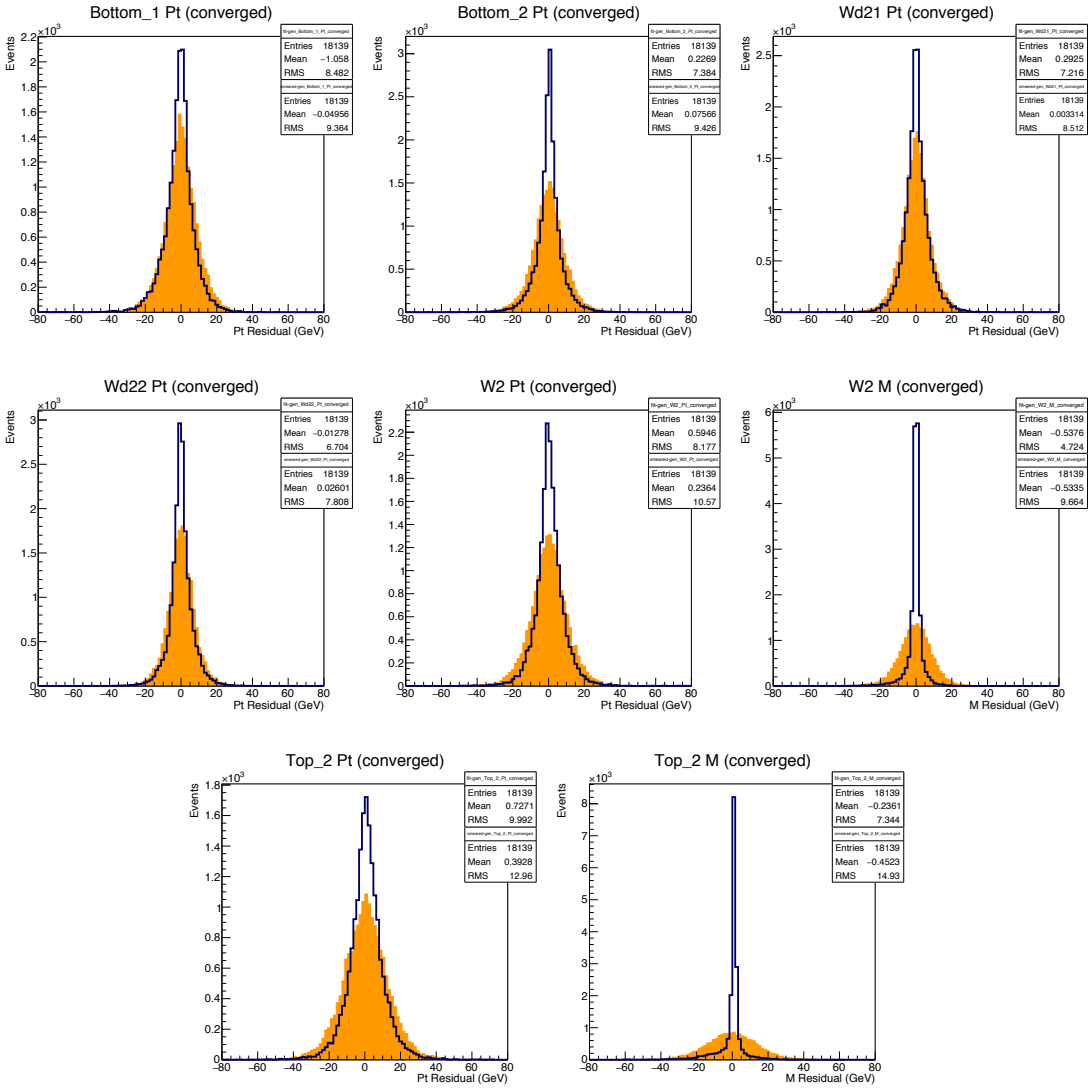


Figure 10.2: p_T values for the b quark from the leptonically-decaying top (Bottom_1), the b quark from the hadronically-decaying top (Bottom_2), the two light quarks from the hadronically-decaying W (Wd21 and Wd22), and the hadronically-decaying W and top (W2 and Top_2), as well as the mass of the hadronically-decaying W and top, for events for which the fit converged. Blue: $p_{\text{fit\_gen}}$; Orange: $p_{\text{smeared\_gen}}$. Wd22 is the W-daughter that we pretended was unmeasurable in the top reconstruction fitting process (see section 7.2).

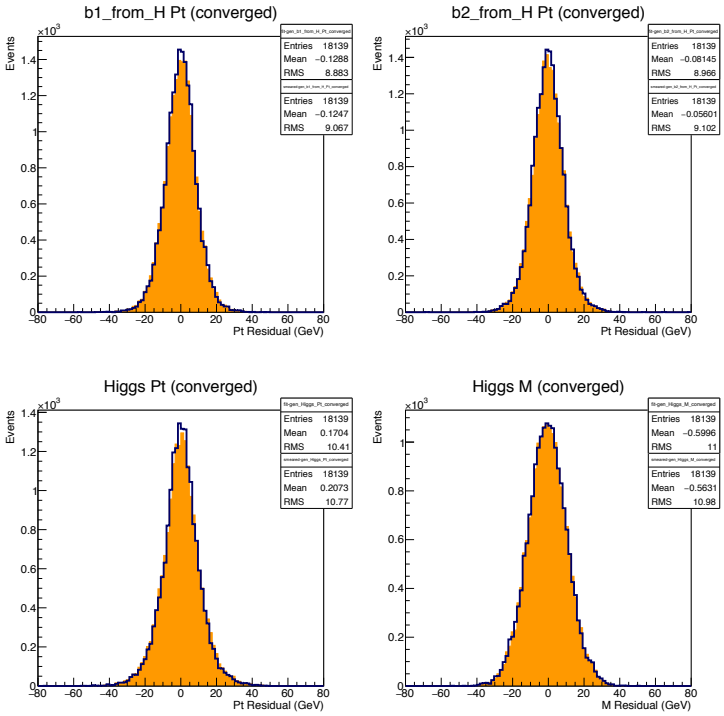


Figure 10.3: p_T values for the non-top system: the two b quark daughters of the Higgs (b1_from_H and b2_from_H) and the Higgs itself, as well as the Higgs mass, for events for which the fit converged. Blue: $p_{\text{fit\_gen}}$; Orange: $p_{\text{smeared\_gen}}$.

10.1.2 Events that Failed to Converge

Figures 10.4, 10.5 and 10.6 contain the distributions of $p_{\text{fit\_gen}}$ and $p_{\text{smeared\_gen}}$, this time for events where the fit failed to converge. At first glance, these plots seem to show a similar trend to the converged events, with the $p_{\text{fit\_gen}}$ distributions looking narrower than the $p_{\text{smeared\_gen}}$ ones. However, inspecting the root mean squared (RMS) values of the distributions yields a different story. For most of the variables, the RMS of the $p_{\text{fit\_gen}}$ distribution is larger than that of the $p_{\text{smeared\_gen}}$ distribution. This shows that there are outlier events in the $p_{\text{fit\_gen}}$ distribution that fall far away from 0, even though this distribution looks narrower on the plot.

As a measure of whether the fitter produced better estimates than the original smeared values, consider the quantity

$$PI = \frac{\text{RMS}(p_{\text{smeared\_gen}}) - \text{RMS}(p_{\text{fit\_gen}})}{\text{RMS}(p_{\text{smeared\_gen}})} \times 100\% \,. \tag{10.3}$$

PI, which stands for "percentage improvement", tells us how much smaller the RMS of $p_{\text{fit\_gen}}$ is compared to that of $p_{\text{smeared\_gen}}$. A positive value of PI indicates that the fit has improved matters; a negative value indicates that the fit has worsened matters.

Table 10.1 shows PI for the different variables shown in the plots, with separate columns for the events which did and did not converge. Note that the RMS value shown in the plot legends are calculated using only the events which fall within the bounds of the histogram in the plot. To better capture outlier events, I therefore used RMS values from histograms with wider ranges than shown in the figures – a range of -5 to 5 for angular variables, and -150 to 150 GeV for variables measured in GeV. We see from the table that PI is positive for all variables in the converged case except for the Higgs mass

(where it takes a very small negative value). On the other hand, for events where the fit failed to converge, PI is negative for most of the variables.

10.1.3 Takeaway

The results of this section show that the top reconstruction fitter is effective, improving our estimates of the particles in the $t\bar{t}$ system, though not much affecting the particles outside that system. However, this result only holds when the fit converges – it would therefore probably be wise to use the output of the kinematic fitter only when it has converged.

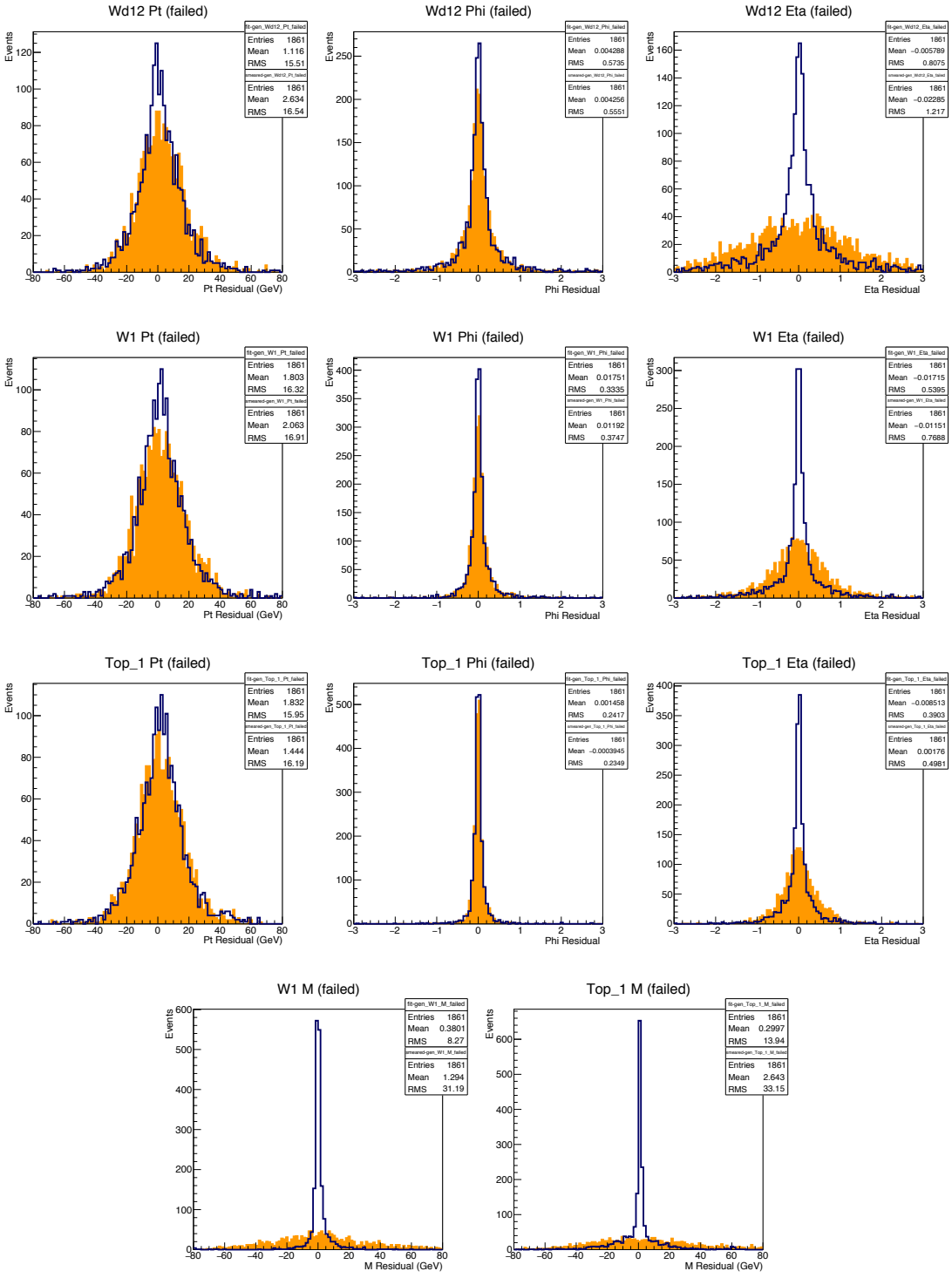


Figure 10.4: p_T, ϕ and η values for the neutrino and leptonically-decaying W and top, as well as W and top mass, for events for which the fit did not converge. Blue: $p_{\text{fit\_gen}}$; Orange: $p_{\text{smeared\_gen}}$. The neutrino is labelled "Wd12", and the leptonically-decaying W and top are labelled "W1" and "Top_1" respectively.

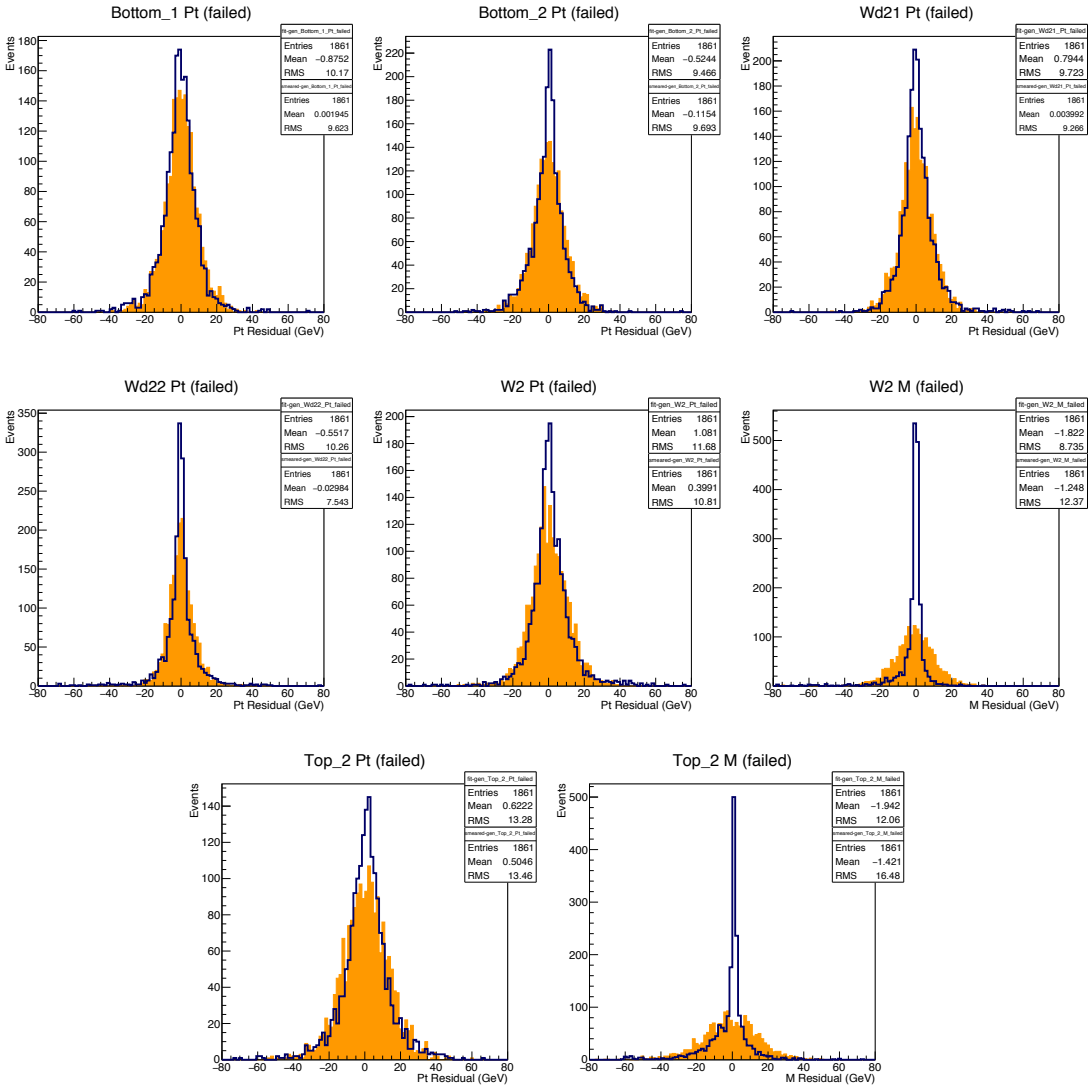


Figure 10.5: p_T values for the b quark from the leptonically-decaying top (Bottom_1), the b quark from the hadronically-decaying top (Bottom_2), the two light quarks from the hadronically-decaying W (Wd21 and Wd22), and the hadronically-decaying W and top (W2 and Top_2), as well as the mass of the hadronically-decaying W and top, for events for which the fit failed to converge. Blue: $p_{\text{fit\_gen}}$; Orange: $p_{\text{smeared\_gen}}$. Wd22 is the W daughter that we pretended was unmeasurable in the top reconstruction fitting process (see section 7.2).

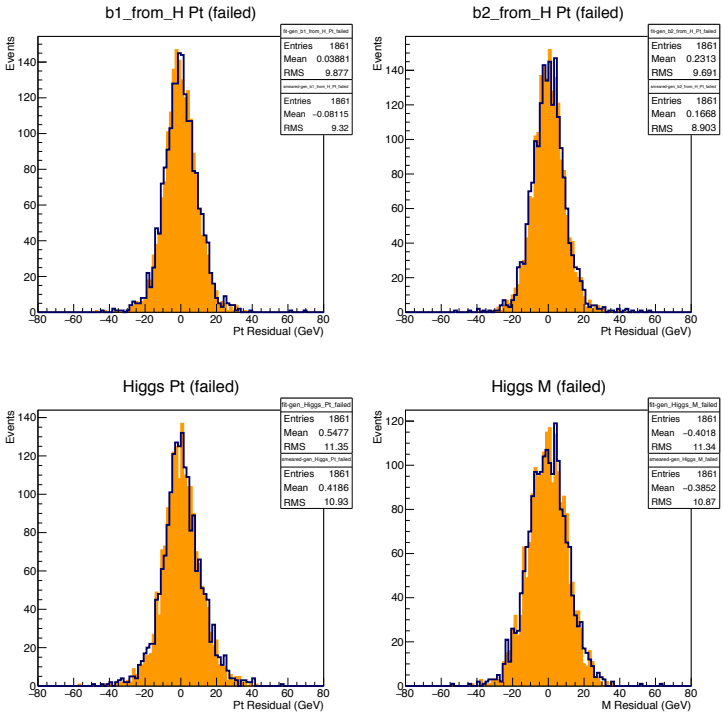


Figure 10.6: p_T values for the non-top system: the two b quark daughters of the Higgs (b1_from_H and b2_from_H) and the Higgs itself, as well as the Higgs mass, for events for which the fit failed to converge. Blue: $p_{\text{fit\_gen}}$; Orange: $p_{\text{smeared\_gen}}$.

Variable	% Improvement in RMS, Converged events	% Improvement in RMS, Failed events
Wd12 Pt	17.0	-2.7
Wd12 Phi	13.3	-1.6
Wd12 Eta	27.4	27.0
W1 Pt	15.6	-4.6
W1 Phi	20.1	10.1
W1 Eta	26.3	27.7
Top_1 Pt	13.4	-3.0
Top_1 Phi	14.0	-6.7
Top_1 Eta	13.1	21.0
W1 M	83.4	71.2
Top_1 M	74.2	66.3
Bottom_1 Pt	6.4	-11.7
Bottom_2 Pt	19.5	-3.7
Wd21 Pt	12.9	-18.5
Wd22 Pt	12.9	-55.0
W2 Pt	20.4	-27.5
W2 M	51.1	30.8
Top_2 Pt	20.3	-19.9
Top_2 M	51.2	29.1
b1_from_H Pt	2.0	-15.3
b2_from_H Pt	1.5	-8.9
Higgs Pt	3.2	-5.6
Higgs M	-0.4	-7.4

Table 10.1: % improvement in RMS (defined in Equation 10.3) of different variables. Negative values (highlighted in red) indicate that $p_{\text{fit\_gen}}$ has a greater spread than $p_{\text{smeared\_gen}}$. RMS values are taken from histograms with wider ranges than those shown in the plots.

10.2 Limit Calculations Using BDT Distributions on Fully-Simulated MC Data

In this section, we apply the top reconstruction fitter to the task of searching for $t\bar{t}H$ events, using the methods described in section 5. We consider the single-leptonic $t\bar{t}H$, $H \to b\bar{b}$ channel, applying the cuts and preprocessing described in section 5.2.1.

10.2.1 Jet and b-tag Multiplicities

Because events with differing numbers of jets and b-jets have different signal and background compositions, it is useful to separate events into regions based on jet and b-tag multiplicities.

For the signal dataset in the single-leptonic case, we expect 6 jets (2 b quarks from the tops, 2 b quarks from the Higgs, and 2 light quarks from the hadronically-decaying W). Because of detector noise and uncertainties in the particle reconstruction algorithms, we do not always get 6 jets. However, the kinematic fitting method requires the presence of 6 jets; we therefore only consider events with at least 6 jets.

Events are then further separated into categories based on b-tag multiplicity. For limit calculation, we consider only two such regions: events with 3 b-tags, and events with 4 or more b-tags (called the 6j3b and 6j4b categories respectively).

10.2.2 Object Permutations

One of the tricky things about the top reconstruction method is that it requires us to know the permutations of quarks in the event – which b quark originated from the hadronic top, leptonic top, and Higgs boson. In an event with 6 jets, exactly 4 of which are b-tagged, there are 12 ways to assign the 4 b-tagged jets to the three parents.

Real events are complicated by a limited b-tag efficiency, so the number of b-tags is usually less than 4 even for a signal MC event. Which jets are b-jets is therefore unclear. If we allow ourselves to permute between all the jets in the event, then an event with 6 jets yields 180 permutations, and an event with more jets would yield even more.

The running time for the reconstruction of one event is about 30 seconds. Evidently, it would take too long to try out even 12 permutations for each event. To get around this problem, we first run a fast permutation algorithm [69], picking out the best 3 permutations, and then applying the top reconstruction fitter to them.

Fast Permutation Calculator

This fast permuter works by calculating, for each possible permutation, two χ^2 measures for each type of parent particle constructed. One measure compares the invariant mass of the daughter particles to the nominal mass of the parent, which is calculated by averaging over the invariant masses of the daughters when using the correct permutation (known in simulated data). The other involves the b-tag CSV value of the daughter jets.

Suppose we have a permutation where we have assigned 6 jets the following roles:

- b_{had}: The b quark from the hadronically-decaying top
- b_{lep}: The b quark from the leptonically-decaying top
- b_{H}: A b quark from the Higgs boson (there are two of these)
- lj: A light jet from the hadronically-decaying W (there are two of these)

We now calculate the χ^2 measures, starting with the one for the hadronically-decaying W:

$$\chi^2_{\text{Whad, mass}} = \frac{\left(m_{\text{inv}}(lj_1, lj_2) - m_{W,\text{had}}\right)^2}{\sigma^2_{m_{W,\text{had}}}}\,, \tag{10.4}$$

where $m_{\text{inv}}(lj_1, lj_2)$ is the invariant mass of the light jets, $m_{W,\text{had}}$ is the nominal W mass, and $\sigma_{m_{W,\text{had}}}$ is its width. We also calculate the measure

$$\chi^2_{\text{Whad, tag}} = \alpha\text{CSV}_{lj_1} + \alpha\text{CSV}_{lj_2}\,, \tag{10.5}$$

where α is a parameter used to set the relative effect of $\chi^2_{\text{Whad, mass}}$ and $\chi^2_{\text{Whad, tag}}$. Here, we set it to 0.6. Notice that we penalise the light jets chosen by this permutation if they have high CSV values, because we expect light jets to have low CSV values.

Next, we calculate the χ^2 measures for the hadronically-decaying top:

$$\chi^2_{\text{thad, mass}} = \frac{\left(m_{\text{inv}}(b_{\text{had}}, lj_1, lj_2) - m_{t,\text{had}}\right)^2}{\sigma^2_{m_{t,\text{had}}}}\,, \tag{10.6}$$

where $m_{\text{inv}}(b_{\text{had}}, lj_1, lj_2)$ is the invariant mass of the b quark and light jets, $m_{t,\text{had}}$ is the nominal top mass, and $\sigma_{m_{t,\text{had}}}$ is its width. We also calculate

$$\chi^2_{\text{thad, tag}} = \alpha(1 - \text{CSV}_{b_{\text{had}}}) + \alpha\text{CSV}_{lj_1} + \alpha\text{CSV}_{lj_2} \tag{10.7}$$

Notice that we penalise b_{had} if it has a low CSV value.

Now let's consider the leptonically-decaying branch. We don't have to bother with the leptonic W, because it doesn't involve any jets. For the leptonic top,

$$\chi^2_{\text{tlep, mass}} = \frac{\left(m_{\text{inv}}(b_{\text{lep}}, \text{lepton}, \nu) - m_{t,\text{lep}}\right)^2}{\sigma^2_{m_{t,\text{lep}}}}\,, \tag{10.8}$$

where $m_{\text{inv}}(b_{\text{lep}}, \text{lepton}, \nu)$ is the invariant mass of the b quark, lepton and neutrino, $m_{t,\text{lep}}$ is the nominal mass of the leptonic top, and $\sigma_{m_{t,\text{lep}}}$ is its width. Since we don't know the neutrino momentum, we substitute the MET and calculate the invariant mass with it. We also calculate

$$\chi^2_{\text{tlep, tag}} = \alpha(1 - \text{CSV}_{b_{\text{lep}}})\,. \tag{10.9}$$

Finally, we consider the particles assigned to the Higgs boson:

$$\chi^2_{\text{H, mass}} = \frac{\left(m_{\text{inv}}(b_{1H}, b_{2H}) - m_H\right)^2}{\sigma^2_{m_H}}\,; \tag{10.10}$$

$$\chi^2_{\text{H, tag}} = \alpha(1 - \text{CSV}_{b_{1H}}) + \alpha(1 - \text{CSV}_{b_{2H}})\,. \tag{10.11}$$

Having calculated all these χ^2 values, we keep the three permutations which have the lowest total χ^2.

Kinematic Fitter χ^2

The three highest-ranking permutations are passed through the top reconstruction fitter. We then consider one further pseudo-permutation: the permutation out of those

three which produced the lowest χ^2 from the top reconstruction kinematic fitter. Of the three permutations, only those for which the fit converged are considered. Which permutation is the best would of course differ for each event.

In the following text, I will refer to the highest, second-highest and third-highest permutations from the quick permutation calculator as permutations 0, 1 and 2 respectively. The permutation out of these three that produced the lowest kinematic fitter χ^2 will be labelled permutation 3.

10.2.3 Signal and Background Processes

The signal and background processes considered in this study, together with their associated MC datasets and their cross-sections, are shown in Tables 10.2 and 10.3.

Process	MC Datasets	$\sigma \times \mathcal{B}$ [pb]
$t\bar{t}H$, $H \to b\bar{b}$	ttHTobb_M125_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8	0.5071 × 0.5824
$t\bar{t}H$, $H \to$ non-$b\bar{b}$	ttHToNonbb_M125_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8	0.5071 × 0.4176

Table 10.2: Signal processes and their MC datasets, cross-sections σ and branching fractions $\mathcal{B}$.

The histograms for each process are normalised by the factor

$$\text{Norm factor} = \frac{35.92\text{fb}^{-1} \cdot \sigma\,[\text{pb}] \cdot \mathcal{B} \cdot 1000}{N}\,, \tag{10.12}$$

where N is the total number of events processed by the code for each process.

Process	MC Datasets	σ [pb]
$t\bar{t}$ + jets	TT_TuneCUETP8M2T4_13TeV-powheg-pythia8	831.76
Single Top	ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1	3.70
	ST_t-channel_antitop_4f_inclusiveDecays_TuneCUETP8M2T4_13TeV-powhegV2-madspin	80.95
	ST_t-channel_top_4f_inclusiveDecays_TuneCUETP8M2T4_13TeV-powhegV2-madspin	136.02
	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4	35.85
	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4	
$t\bar{t} + W$	TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
$t\bar{t} + Z$	TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.2529
	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297
W + jets	WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	3091.52
Diboson	WW_TuneCUETP8M1_13TeV-pythia8	118.7
	WZ_TuneCUETP8M1_13TeV-pythia8	47.12
	ZZ_TuneCUETP8M1_13TeV-pythia8	31.73

Table 10.3: Background processes and their MC datasets and cross-sections σ.

10.2.4 Systematic Uncertainties

For simplicity, we used only those systematic uncertainties which affect the normalisation, but not the shape, of the histograms. The uncertainties for each process are shown in table 10.4. In addition, the uncertainty in integrated luminosity, which affects all processes equally, is 2.5 %. All the uncertainties used are taken to have a log-normal distribution.

Process	pdf				QCD Scale				
	$gg_{\mathrm{t\bar{t}H}}$	gg	$q\bar{q}$	qg	$t\bar{t}$	t	V	VV	$\mathrm{t\bar{t}H}$
$\mathrm{t\bar{t}H}$	3.6 %								-9.2%/+5.8%
$\mathrm{t\bar{t}}$ + jets		4 %			-4%/+2%				
Single Top				3%		-2%/+3%			
$\mathrm{t\bar{t}} + W$			2%		-12%/+13%				
$\mathrm{t\bar{t}} + Z$		3%			-12%/+10%				
W + jets			4%				1%		
Diboson			2%					2%	

Table 10.4: Systematic uncertainties used.

10.2.5 BDT Discriminator Distributions

The BDT discriminator is calculated (as described in section 5.3.1) for each permutation of the events passed through the kinematic fitter, as well as for the original events which were not kinematically fitted. The results of section 10.1.2 suggest that we should use the results of the top reconstruction fitter only if the fit converges. Thus, for non-converging permutations, the BDT discriminator value of the original vanilla particles (the ones not passed through the fit) is used instead.

The BDT discriminator calculation is based on BDTs that were previously trained on the vanilla events. Since the training was done on odd-numbered events, we only use even-numbered events for the distributions, in order to avoid bias.

The missing transverse energy is one of the inputs to the BDT for some tag multiplicity categories. The value of the MET depends on the momenta of all the particles involved in the event. For the kinematically reconstructed events, we therefore re-calculate the MET based on the new momenta values of the visible particles.

Figures 10.7 through 10.13 show the BDT distributions for vanilla and kinematically-reconstructed events, for various processes and categories.

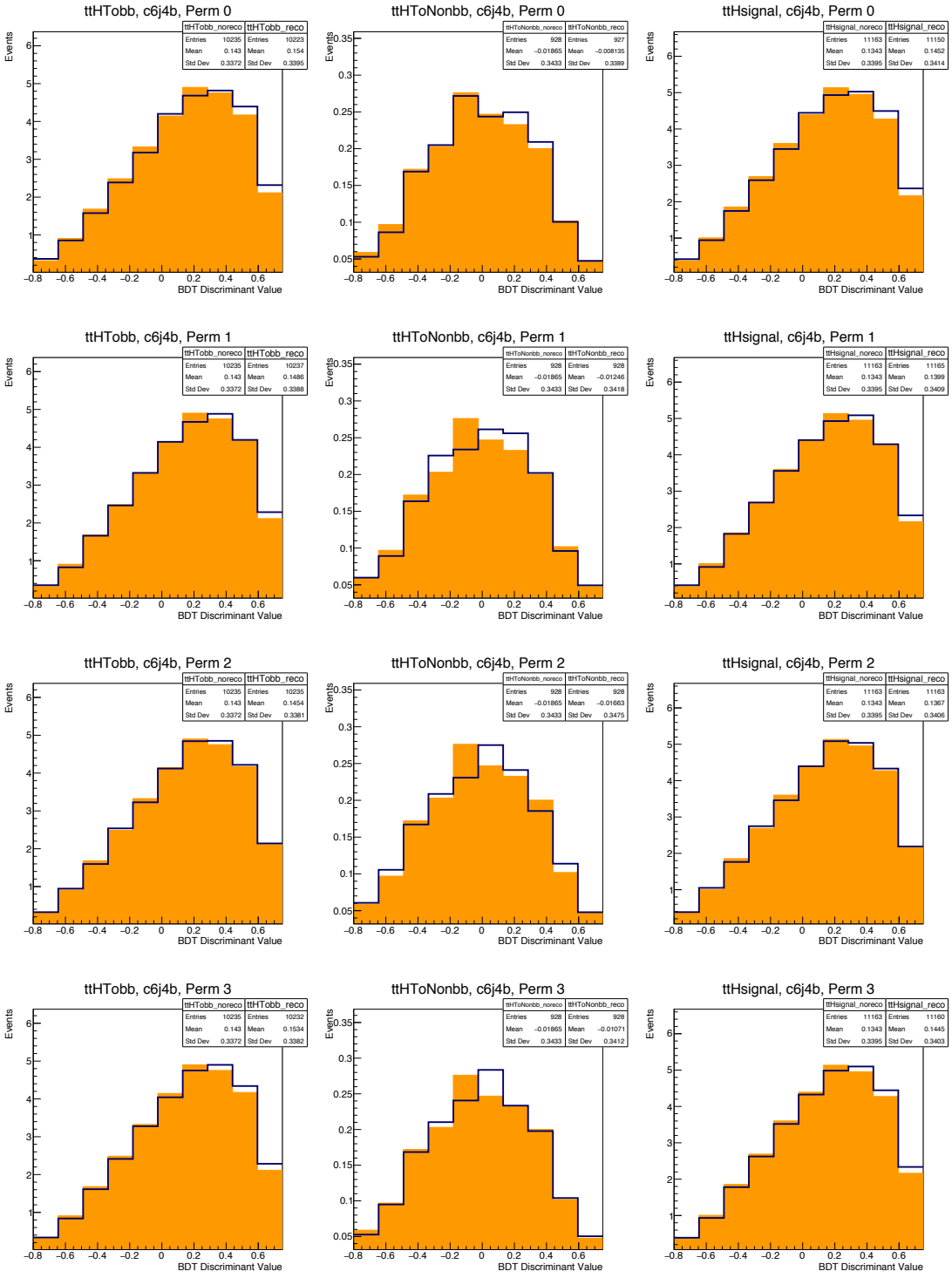


Figure 10.7: BDT discriminant values for signal events in the 6j4b category. Left column: $t\bar{t}H$, $H \to b\bar{b}$; Middle column: $t\bar{t}H$, $H \to$ non-$b\bar{b}$; Right column: both processes combined. Orange: vanilla events; Blue: distribution after kinematic reconstruction. The four rows correspond to the four permutations described in section 10.2.2.

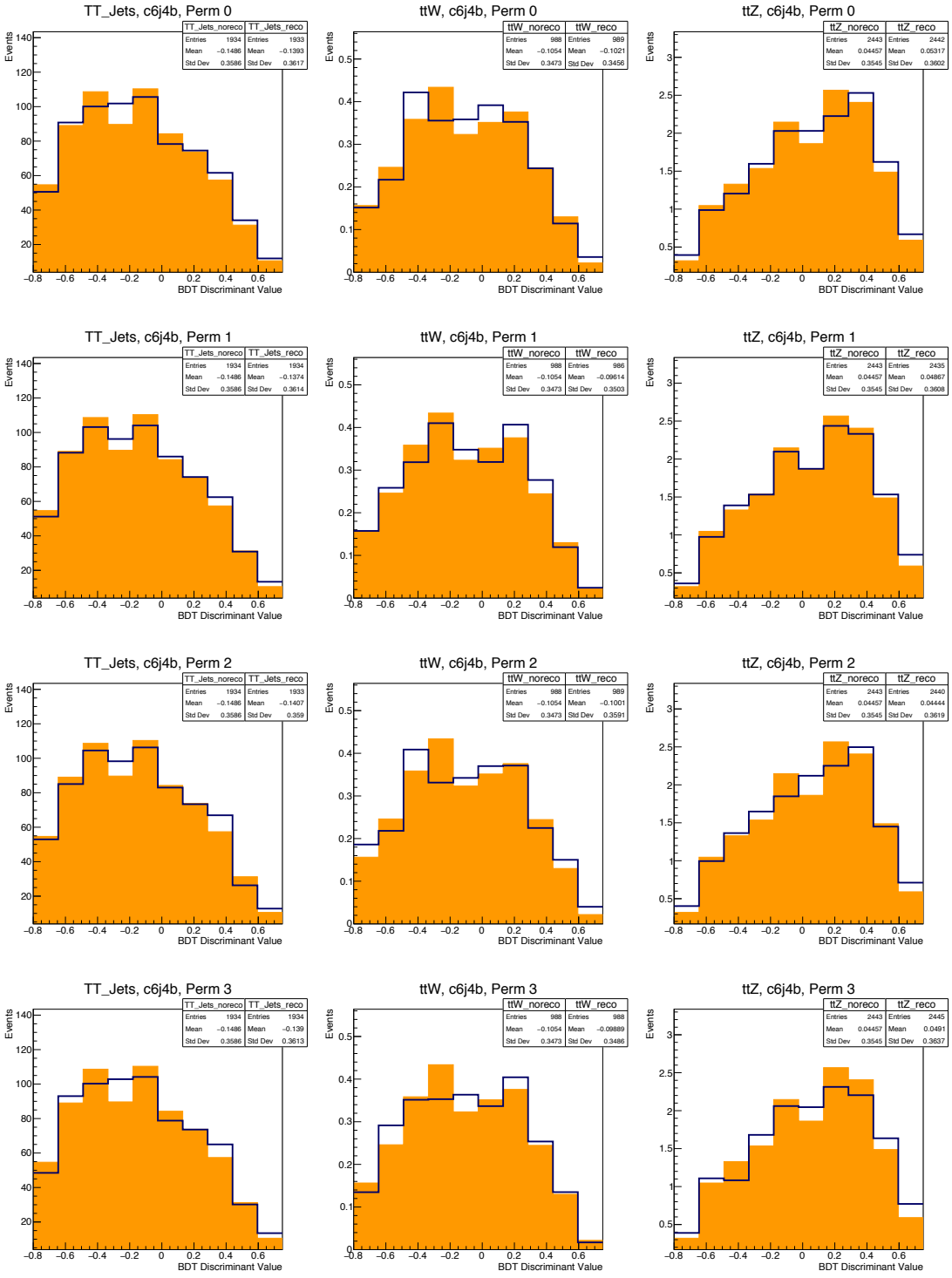


Figure 10.8: BDT discriminant values for some of the background events in the 6j4b category. Left column: $t\bar{t}$+ jets; Middle column: $t\bar{t} + W$; Right column: $t\bar{t} + Z$. Orange: vanilla events; Blue: distribution after kinematic reconstruction. The four rows correspond to the four permutations described in section 10.2.2.

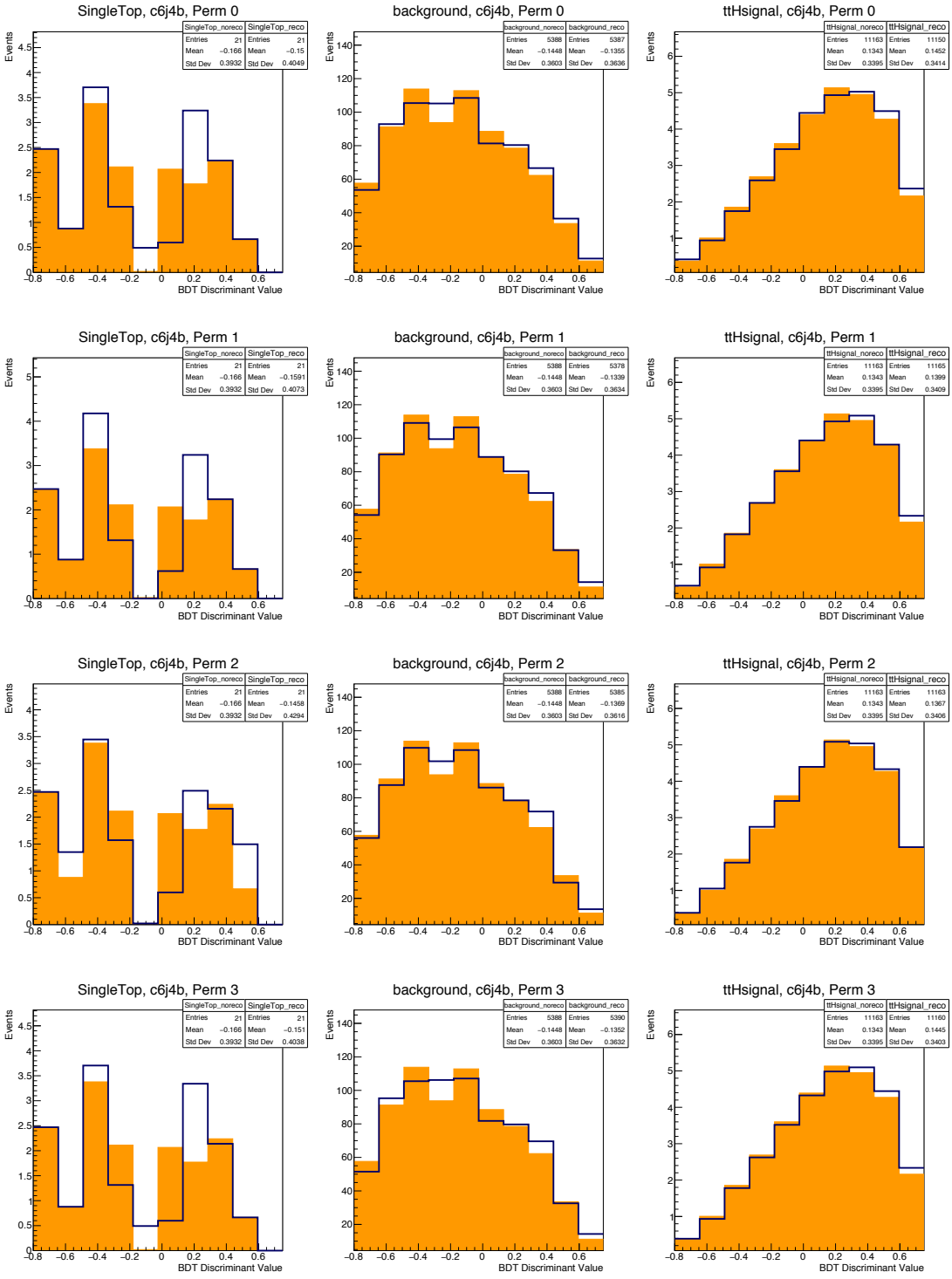


Figure 10.9: BDT discriminant values in the 6j4b category, for the single-top background (left column), as well as the combined background (middle column), with the combined signal distribution (right column) included for comparison. The diboson and $W+$ jets processes produced zero yield in this category, so their plots are not shown here. Orange: vanilla events; Blue: after kinematic reconstruction. The four rows correspond to the four permutations described in section 10.2.2.

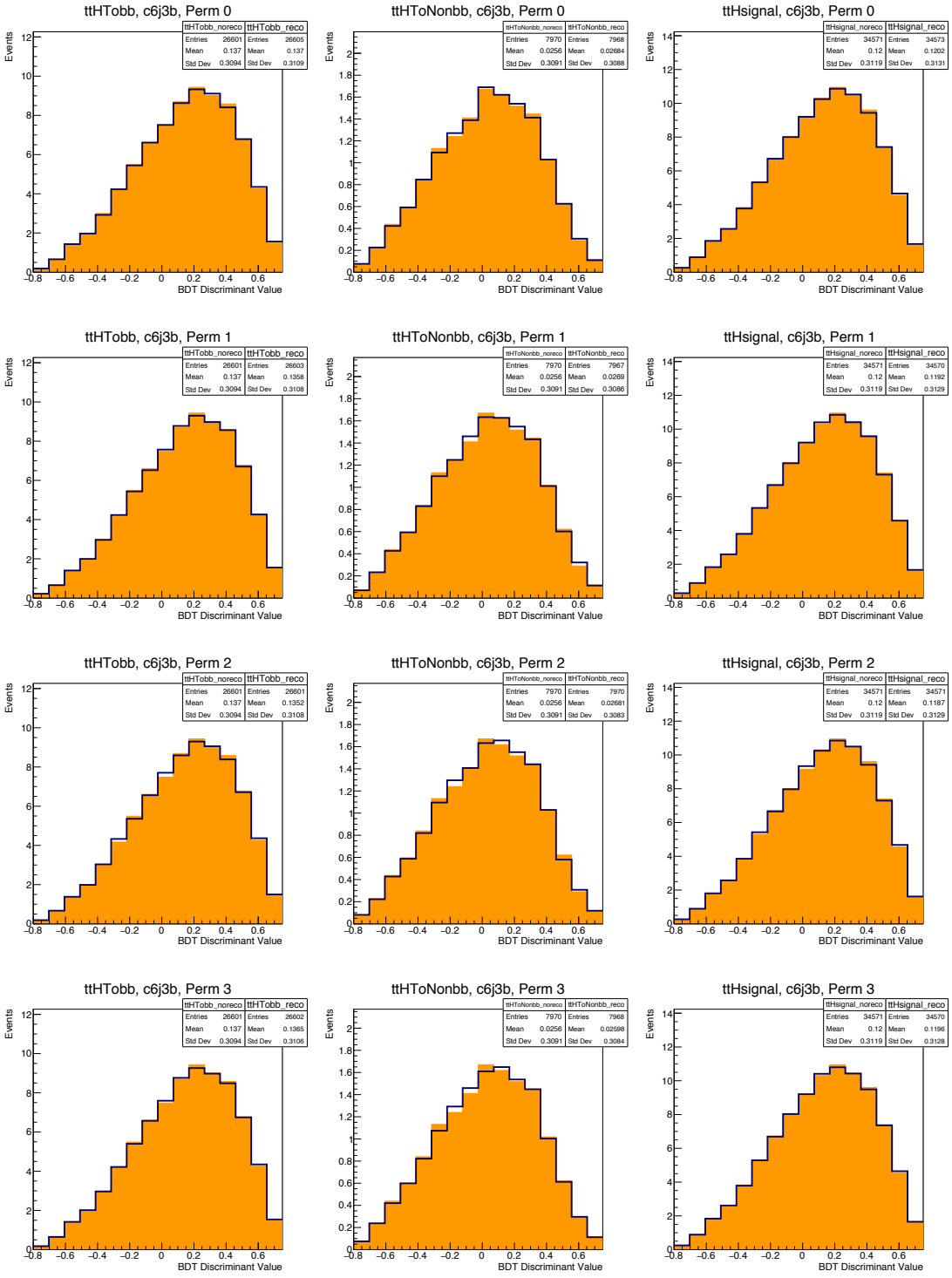


Figure 10.10: BDT discriminant values for signal events in the 6j3b category. Left column: $t\bar{t}H$, $H \to b\bar{b}$; Middle column: $t\bar{t}H$, $H \to$ non-$b\bar{b}$; Right column: both processes combined. Orange: vanilla events; Blue: distribution after kinematic reconstruction. The four rows correspond to the four permutations described in section 10.2.2.

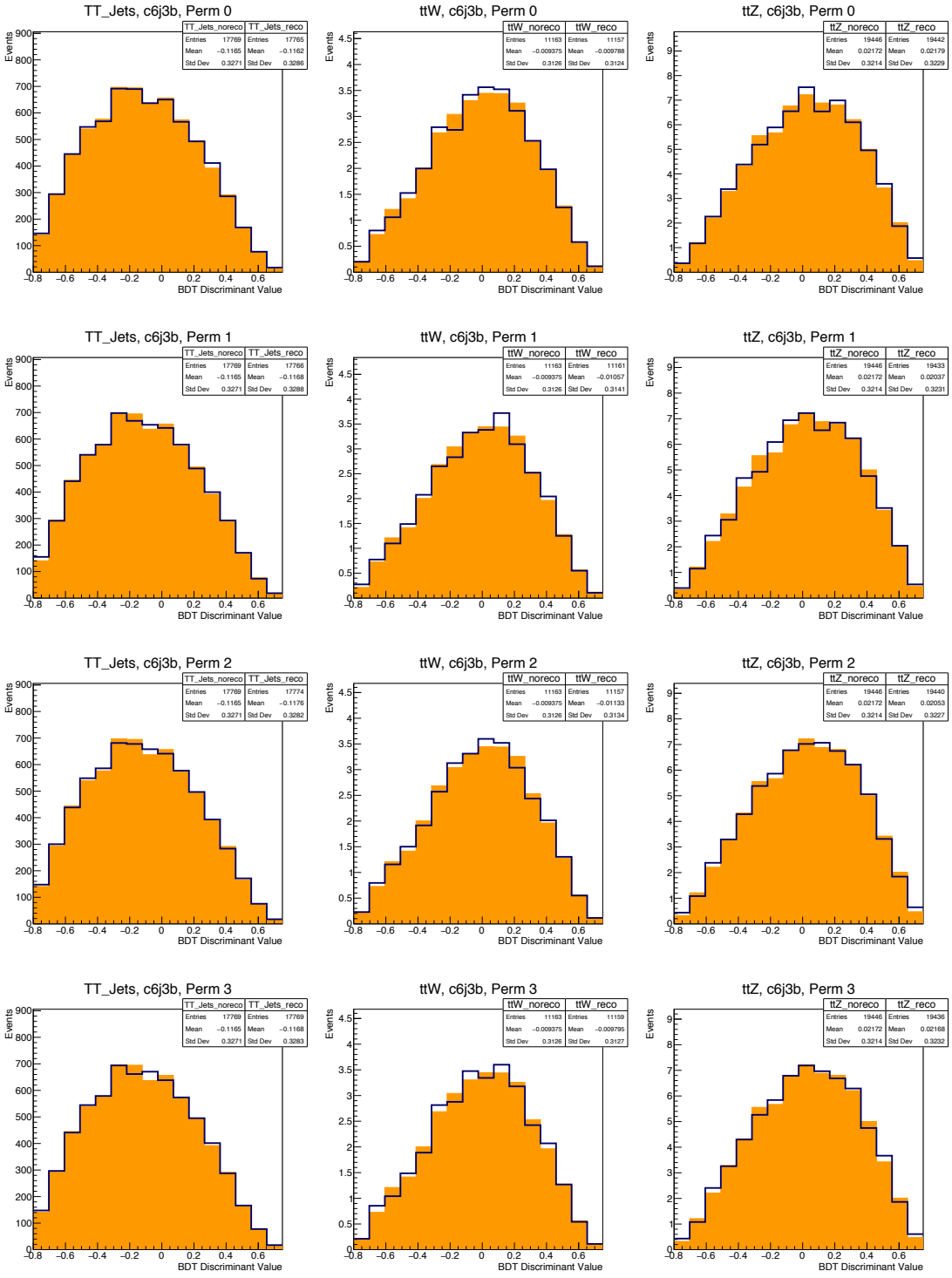


Figure 10.11: BDT discriminant values for some of the background events in the 6j3b category. Left column: $t\bar{t}$+ jets; Middle column: $t\bar{t} + W$; Right column: $t\bar{t} + Z$. Orange: vanilla events; Blue: distribution after kinematic reconstruction. The four rows correspond to the four permutations described in section 10.2.2.

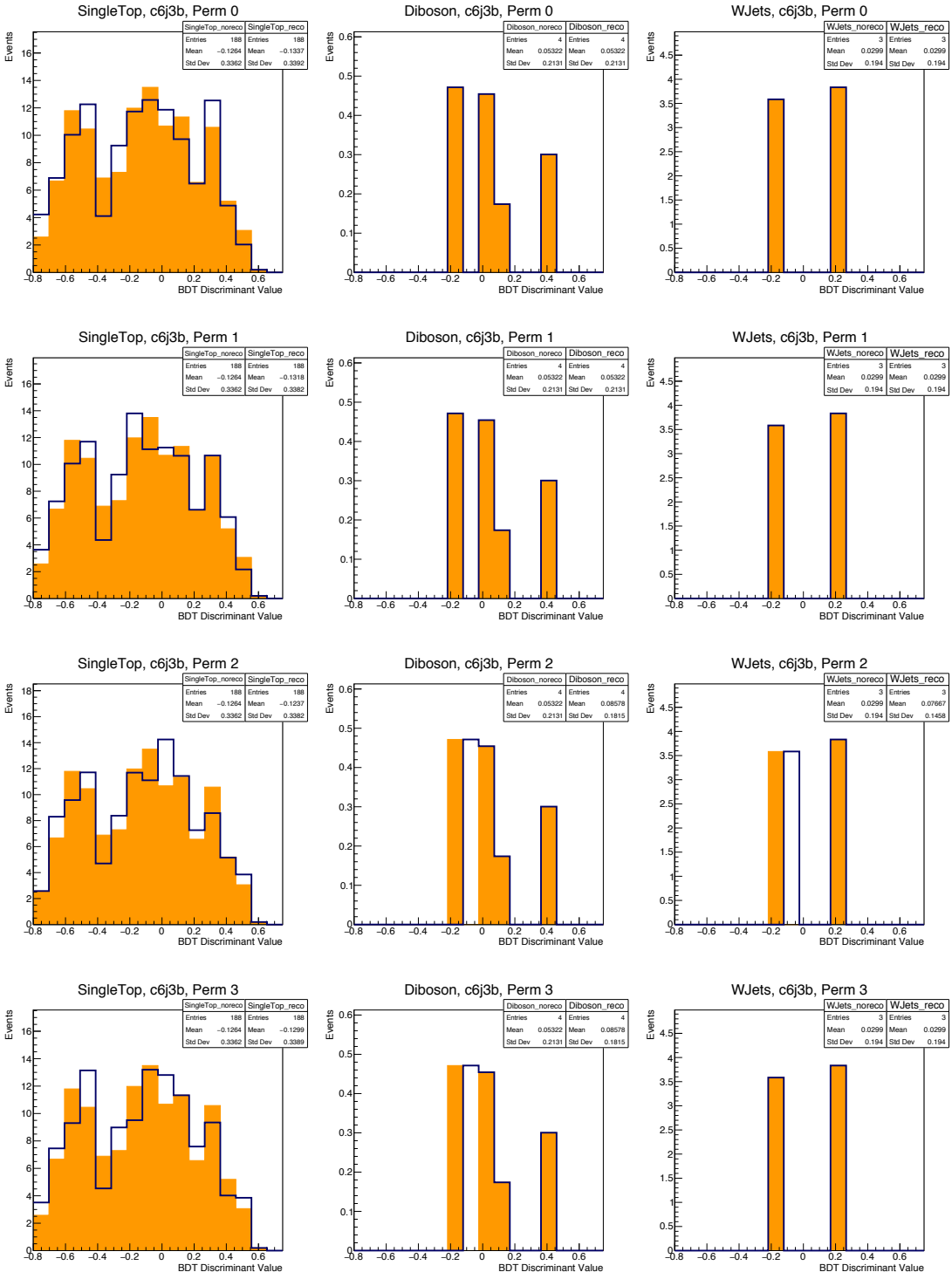


Figure 10.12: BDT discriminant values for some of the background events in the 6j3b category. Left column: single-top; Middle column: diboson; Right column: W + jets. Orange: vanilla events; Blue: distribution after kinematic reconstruction. The four rows correspond to the four permutations described in section 10.2.2.

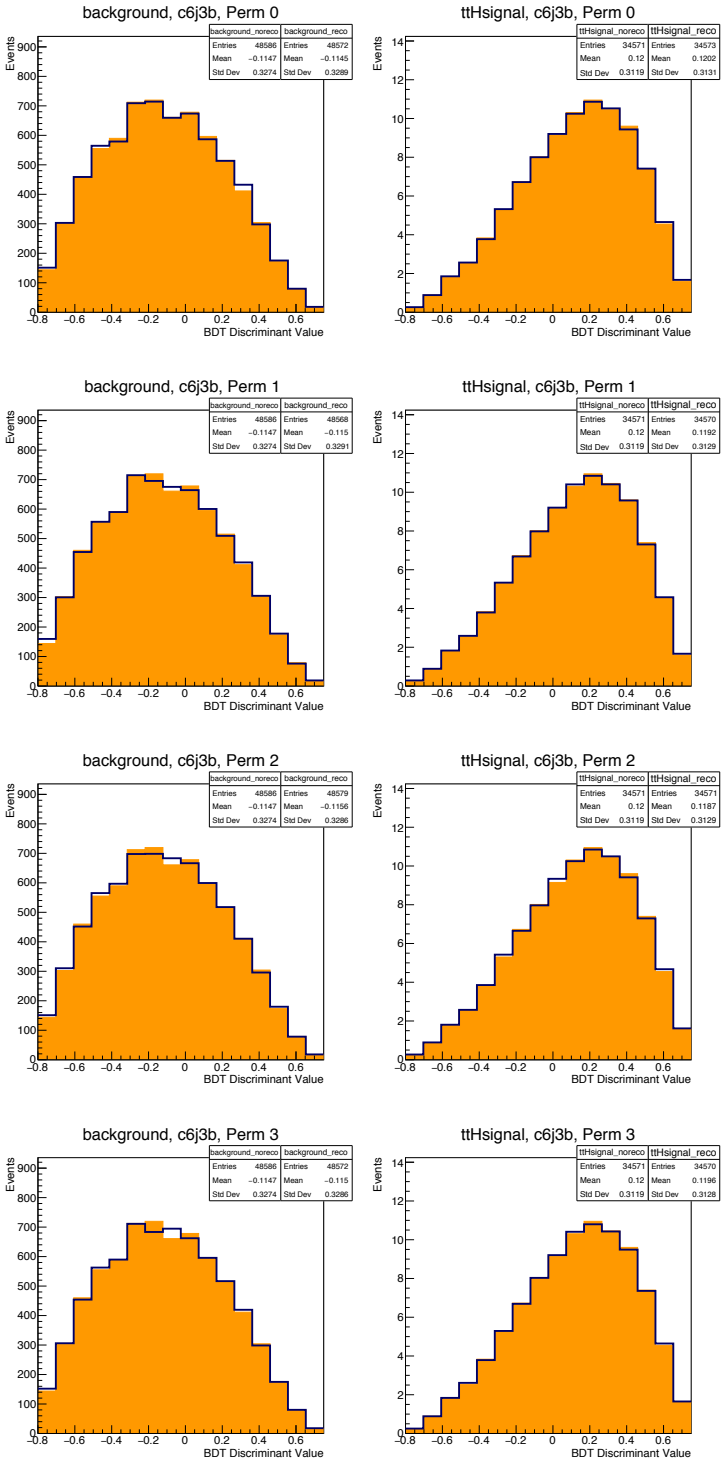


Figure 10.13: BDT discriminant values in the 6j3b category, for the combined background (left column), with the combined signal (right column) included for comparison. Orange: vanilla events; Blue: distribution after kinematic reconstruction. The four rows correspond to the four permutations described in section 10.2.2.

At first glance, the BDT distributions for the signal events (Figure 10.7) for the $t\bar{t}H$, $H \to b\bar{b}$ process in the 6j4b category seem promising. The kinematically-reconstructed distribution is shifted visibly to the right compared to the vanilla one, indicating that these events are seen as more signal-like after the reconstruction. The effect is strongest in Perm 0, and gets increasingly weaker as we move to permutations 1 and 2, as we would expect. Surprisingly, Perm 0 shows a slightly greater rightward shift in the mean than Perm 3 (recall that the latter consists of the best permutation for each event, as measured by the χ^2 value output by the kinematic fitter).

The $t\bar{t}H$, $H \to$ non-$b\bar{b}$ process does not show a strong rightward shift after the reconstruction, but since these events are heavily outnumbered in the 6j4b category by the $t\bar{t}H$, $H \to b\bar{b}$ process, the overall signal distribution is quite similar to the distribution of the latter process. In the combined signal, the mean of the BDT distribution shifts rightwards by 8.1% after the reconstruction for Perm 0, compared to 7.6% for Perm 3.

The picture becomes less rosy, however, when we study the background distributions (Figures 10.8 and 10.9). In the 6j4b category, the largest background, $t\bar{t}$+ jets, also shows a visible rightward shift of the BDT distribution after reconstruction – though this shift is not quite as pronounced or clean-cut as for the signal. This effect is also quite strong in the $t\bar{t} + Z$ distribution, but less so for the $t\bar{t} + W$ process. The non-$t\bar{t}$ backgrounds, in the meantime, have too low an occupancy in the 6j4b category to discern the effect of the reconstruction. The shift in the mean for the combined background distribution is about 6.4 % for Perm 0 and 6.6 % for Perm 3, only slightly less than that of the signal distribution.

The 6j3b category, on the other hand, shows little change in the BDT distributions before and after the reconstruction, for both signal and background (Figures 10.10

through 10.13).

10.2.6 Expected Limits

The expected limits at 95% C.L. on the $t\bar{t}H$ signal strength, calculated using the vanilla BDT distributions as well as the distributions for each permutation of the reconstruction, are shown in Figure 10.14 and Table 10.5. There are three sets of limits: calculated using just the 6j4b category, just the 6j3b category, and both categories combined. The "signal-injected" limit is calculated by making a set of "fake data" by taking the sum of counts over all the processes, each normalised as in Equation 10.12.

	6j3b		6j4b		6j3b + 6j4b	
	Signal-Injected	Median	Signal-Injected	Median	Signal-Injected	Median
Perm 3	3.0372	2.2188	2.7805	1.9062	2.2548	1.3789
Perm 2	3.0420	2.2188	2.7819	1.9062	2.2602	1.3789
Perm 1	3.0442	2.2266	2.8133	1.9453	2.2716	1.3945
Perm 0	3.0363	2.2109	2.7755	1.8984	2.2530	1.3750
Vanilla	3.0274	2.2109	2.7722	1.8984	2.2493	1.3711

Table 10.5: Expected limits on the $t\bar{t}H$ signal strength, calculated using the vanilla BDT distributions as well as the distributions for each permutation of the reconstruction.

We see that the reconstruction has made very little difference in the limits, for both categories. Based on the BDT distribution histograms shown in section 10.2.5, this lack of change for the 6j3b category is likely due to the fact that the BDT distributions themselves have not been much affected by the reconstruction process. For the 6j4b category, on the other hand, it seems that both signal and background were almost equally affected by the reconstruction, such that the end effect on the limits was negligible.

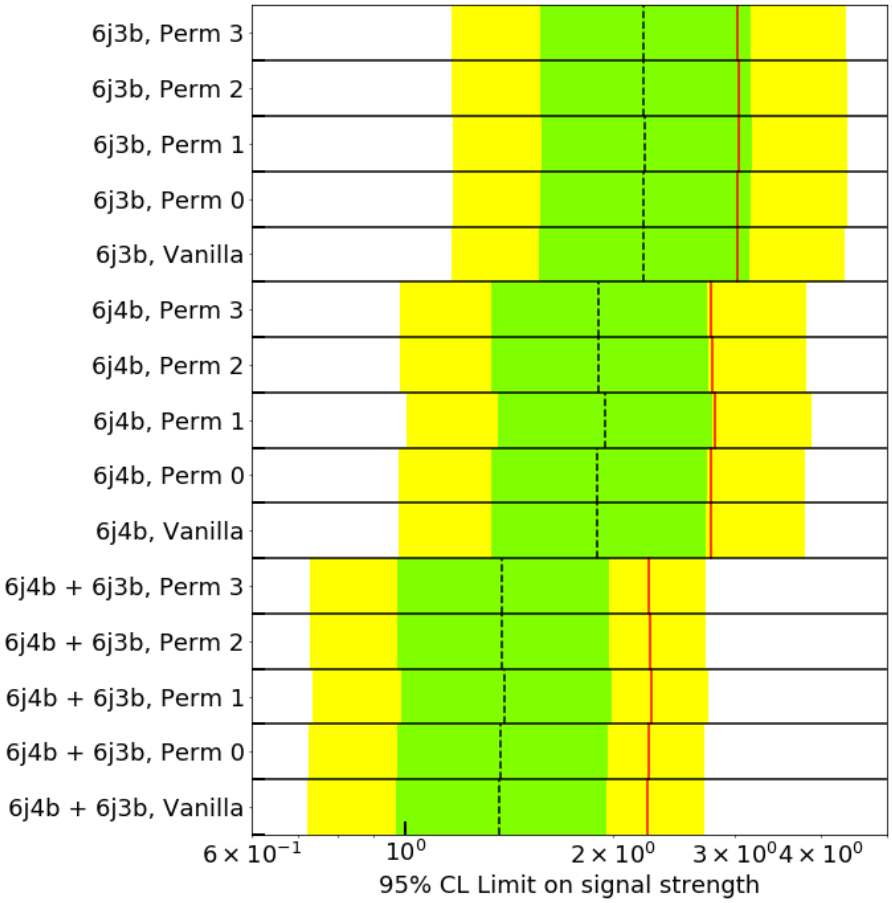


Figure 10.14: Expected limits on the $t\bar{t}H$ signal strength, calculated using the vanilla BDT distributions as well as the distributions for each permutation of the reconstruction. The three sets of limits are calculated using just the 6j4b category, just the 6j3b category, and both categories combined. Black dashed line: median expected limit; Green band: $\pm 1\sigma$; Yellow band: $\pm 2\sigma$; Red line: signal-injected.

10.2.7 Discussion

Since the effect of the kinematic reconstruction is to improve our estimates of particle momenta, we might expect that the reconstruction would allow us to better distinguish between signal and background events, thus improving the limits set on the signal process. However, the results just presented show little change in the calculated limits.

Permutations

One possible reason for this disappointing performance could be the uncertainty regarding b-quark permutations. The results of Section 10.1 show that the reconstruction results in improved estimates of the $t\bar{t}$ system momenta when the correct permutation is fed to the fitter. However, in a simulation of real data, this correct permutation is not known. The quick permutation calculator described in section 10.2.2 can give us only a guess at the best permutations, and in fact produces the correct permutation (as its top-ranking permutation) only 37% of the time when applied to $t\bar{t}H$, $H \to b\bar{b}$ processes [69]. By considering the top three permutations produced by this calculator, and choosing the best of them (Perm 3) based on the kinematic reconstructor's χ^2 output, we might hope to capture the correct permutation and thus reap the full effects of the kinematic reconstructor.

We observe, however, from the BDT distributions and calculated limits, that Perm 3 does not perform better than Perm 0, and in fact sometimes performs slightly worse. This indicates either that the correct permutation is not often found in the top three permutations produced by the quick calculator, or that the kinematic reconstructor χ^2 does not distinguish the correct permutation reliably. On the other hand, the fact that the signal BDT distribution is more right-shifted for Perm 0 compared to Perm 1 and 2 does tend to indicate that Perm 0 is more often the correct permutation than the lower-ranking ones.

Another thing to note is that the permutation calculator assumes the presence of a Higgs boson which decays into two b-quarks (based on its use of the jets' CSV values in calculating its χ^2 measures). Thus, we should in fact only trust it in the case of the $t\bar{t}H$, $H \to b\bar{b}$ process. It is somewhat surprising that the distributions of the $t\bar{t}H$, $H \to$ non-$b\bar{b}$

process and of the background processes were not adversely affected after the application of the permutation calculator and kinematic reconstructor.

Sensitivity to the $t\bar{t}$ System

The BDT distributions of the 6j4b category show, in general, a rightward shift after reconstruction, while the distributions in the 6j3b category are mostly unaffected. Why should the two categories react differently to the reconstruction process? The quick permutation calculator is likely not the cause of this difference. Since the permutator takes as input the raw CSV discriminant values of the jets, rather than their discrete b-tag labels, it is unlikely that it should behave drastically differently when there are 4 b-tags instead of 3.

The BDT discriminant calculation, however, is a different story. The BDTs are trained separately for each category, so the 6j4b category uses a different tree structure and set of input variables from the 6j3b category.

The results in section 10.1 indicate that the reconstruction has an improving effect on the $t\bar{t}$ system, but does not affect the non-top objects much. It would therefore seem that the BDTs in the 6j4b category are more sensitive to the $t\bar{t}$ system than those in the 6j3b category.

Re-Training the BDTs

For both the vanilla and reconstructed events, the BDT discriminants were calculated using the same BDTs, trained on the vanilla datasets. The above observations suggest that the BDTs ought to be re-trained on the reconstructed events, in order to make them

more sensitive to the effect of the reconstruction.

The re-training could be done on several levels. We could maintain the hyper-parameters defining the tree structures and the boosting process, such as the tree size and shrinkage. We could also leave the set of input parameters the same as in the vanilla case. The only parameters to be re-trained would then be the variable used in each node, the splitting point of each node, and the weight of each tree – all this can be trained using the stochastic gradient boosting process described in section 5.3.1. This way of re-training would allow the new BDTs to adjust to the change in the distributions of the input variables, brought about by the reconstruction.

The reconstruction process does give us one additional variable that was unknown in the vanilla events – the neutrino z-momentum. In addition, the application of the permutation calculator also assigns a "role" to each jet (i.e. identifies its parent). This additional information, which was not available to the original BDTs, could be included as possible inputs to the new BDTs. In order to properly incorporate the new potential inputs, we would have to carry out a more extensive re-training process, which includes re-running the particle swarm optimisation algorithm described in section 5.3.1. This process allows us to pick out the best input variables and tree structure for each category. This more extensive re-training process would naturally be more resource-intensive and time-consuming than simply re-running the stochastic gradient boost algorithm.

Given that the reconstruction process produces several permutation possibilities, each of which would have to be separately re-trained, the extensive re-training process might not be feasible. However, note that the main purpose of the extensive re-training would be to incorporate the new input variables generated by the reconstruction, and to pick the best ones. Since each permutation of the reconstruction produces the same set of

input variables, it may be necessary to run the particle swarm algorithm only once on one of the permutations. The resulting tree structure and set of input variables can then be used for all permutations of the reconstructed events. Each permutation can then be separately trained using stochastic gradient boosting.

These newly-trained BDTs, with their advantages of a greater choice of input variables and a better understanding of the post-reconstruction variable distributions, would have the potential to allow a better discrimination between signal and background after the reconstruction, hence improving the limits calculated.

CHAPTER 11

AT THE LARGEST MACHINE IN THE WORLD, PART II: OUTREACH AT CERN

You know, this conversation with you has really inspired me. At CERN, I get these emails sometimes asking for people to volunteer at outreach events, and I think I'm gonna do that.

Oh, that's great! What sort of outreach events does CERN have?

There are guided tours for the public, and also we have a lot of school groups visit. They usually get tours, and can also do activities in S'Cool Lab, which is a laboratory where they do experiments like building cloud chambers using dry ice and alcohol, or playing with an electron tube.

How about the tours? Do visitors actually get to see the LHC and the detectors?

Well, you can see them during the shutdown periods, but the majority of the time the machine is running, so you can't get to it. But there are some older, above-ground machines which are on the tour itineraries.

CERN welcomes more than 100 000 visitors per year from around the globe. These include the general public, who can sign up to join daily guided tours on the CERN website. School groups and private groups (such as companies) also make up a large percentage of the visitors.

There are several fixed itineraries that visiting groups follow. These may involve decommissioned machines, active above-ground machines, and, during shutdown periods, the underground detectors of the LHC. The CERN Visits Service works with the safety officers and other personnel who work at the relevant locations, to make these areas suitable for visits without impeding the work done there. At many locations, the visits service has installed models, posters, videos and viewing platforms for the benefit of the visitors. Visit groups are accompanied by tour guides, who must stick to a predetermined schedule and route for logistics and safety reasons.

CERN also has two visitor centres which the public can explore, without the need to be accompanied by a tour guide. These are the Microcosm visitor centre near the reception building, and the Universe of Particles exhibition in the wooden Globe across the street.

As a tour guide, I mainly led itineraries to the Synchrocyclotron (SC), the ATLAS control room and visitor centre (AVC), and the Antiproton Decelerator (AD). The synchrocyclotron (Figure 11.1) was CERN's first accelerator, built in 1957 and decommissioned in 1990. It has since become a sort of museum, complete with an exhibition of contemporary Geiger counters and mathematical drawing tools, and a snazzy light show that plays on the surface of the machine itself. The ATLAS control room provides a literal window into the lives of CERN scientists, whom visitors can watch sitting in front of screens covered with visualisations of the detector. The adjoining visitor centre contains models of the ATLAS detector and a 3D-video viewing room. Finally, the Antiproton Decelerator is an active experiment which aims to study antiprotons produced by the LHC. It decelerates antiprotons to manageable speeds, before trapping them to form antihydrogen atoms, or performing other experiments with them.

The SC and AVC are usually combined into one 2-hour itinerary for the visitors from the general public. The AD, on the other hand, is usually reserved for private groups or school groups.

Apart from acting as a tour guide, I also facilitated experiments at S'Cool Lab.

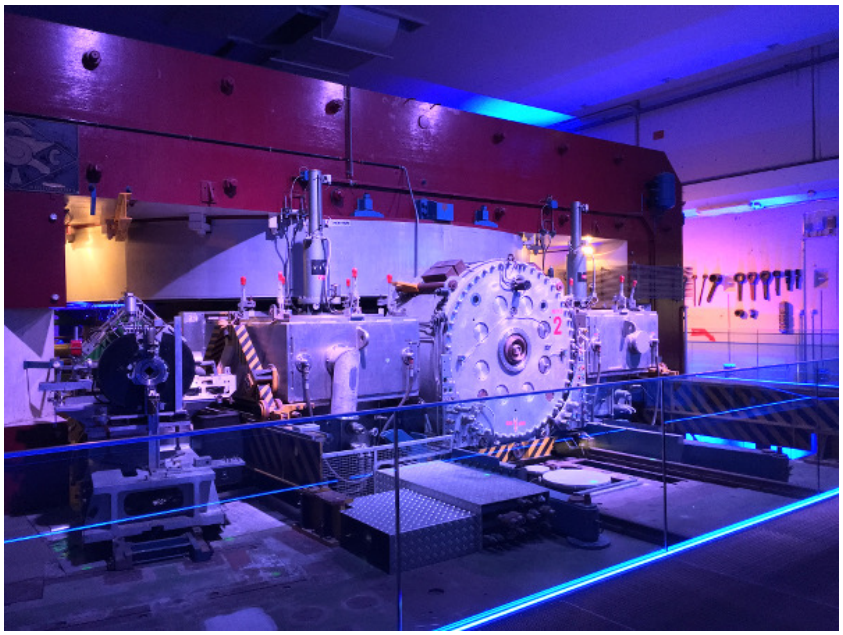


Figure 11.1: The Synchrocyclotron, bathed in a mysterious blue glow.

One challenge with guiding tour groups at CERN is that there really isn't very much to see. The LHC is usually out of bounds, and even smaller machines like the AD are hidden behind large concrete blocks, so that visitors only see the offices and stacks of computing devices located in the same building. Even when the apparatus is visible, as in the case of the SC, the machine is still static, and visitors have to imagine what used to go on inside (with the help of animations from the light show). While the models and videos are very helpful, they are not the "real thing". The commentary provided by the tour guide is thus of utmost importance in making the visit interesting for the participants.

To gather material to talk about, I referred to resources on the CERN document server for guides, which provided information about the various machines. While some of this information was overly technical, there were often useful tidbits and numbers that I otherwise would not have known about. In addition, I would note down and re-use effective analogies that I heard from other outreach practitioners at CERN. I also applied different techniques to construct commentaries that would be relevant and interesting to the audience.

11.1 Leading an Interesting CERN Tour

11.1.1 Providing Background

So, maybe I'll take advantage of you being here to ask you for some tips about how to give a CERN tour.

Sure. Why don't you start by telling me how *you* would start off a tour?

Um, well, I'd explain what CERN does.

Well, if it were me, I'd first introduce myself and ask everyone where they're from, just to build a bit of rapport :). But anyway, you were saying you would explain what CERN does...

Yeah -- we collide particles and... look at what comes out of the collisions. And from there, we can figure out what the universe is

made of.

That was pretty good layman language. But it was a little vague. You can go into more detail; you do have 2 hours, after all. And you could try to make it a bit less abstract. Maybe something like this:

So let's have a quick crash-course in particle physics! Let's take something small, say a human hair. What happens if you zoom in on that, what would you see? (A: Cells, proteins, etc.) And if you zoom in on that? (Molecules) And if you zoom in on a molecule? (Atoms) And if you zoom in on atoms? (Protons, electrons etc.) And if you zoom in on a proton? (Quarks) And zooming in on a quark? (...) That's a trick question – there's nothing. Quarks and electrons are fundamental particles, which means that (as far as we know right now) they're not composed of smaller elements.

And now that you've established your basic particle zoo, you can go on to explain the Standard Model table, and how you can only see the heavier particles in a collider.

Ahah, and I can explain that we make heavy particles by turning energy into mass, from the equation $E = mc^2$. No wait, I'm not supposed to use equations...

Oh, but you can use that one! You're right that you should usually avoid equations, but this one is the most famous equation in science – everybody knows it. In fact, even better – ask people to tell you what each symbol means. People who are interested enough in science to sign up for a CERN tour will know this, and it will make them feel good to be able to answer your question.

11.1.2 Analogies

Particle physics is very abstract. Not only does it deal with things invisible to the naked eye, the numbers involved are often on a scale that people can't get a physical feeling for. Using analogies and comparisons helps to relate these concepts to everyday experience. Some of my favourite analogies include:

- Likening the CMS detector to a 3-dimensional camera – a 66-megapixel one. That's not that many megapixels, considering that a modern smartphone camera has about 12 megapixels. So what makes CMS so special? It can take 40 million pictures per second!
- Why did CERN have to build increasingly larger accelerators as the particle energies increased? Imagine driving a car and going round a bend – if you get to a really sharp turn, you have to slow down quite a lot. If the turn is more shallow, you can go round at a faster speed.
- The LHC produces 15 million Gigabytes of data per year. That's really difficult to visualise, so instead try this: it's equivalent to a stack of CDs 20 km tall. (That's higher than most planes fly.)
- CERN's annual budget is about 1 billion euros. That sounds like a lot, until you equate it to the cost of one cup of coffee per European Union resident.

11.1.3 The Wow Factor

As a famous experiment which aims to search for the fundamental truths of the universe while pushing the frontiers of technology, the LHC is perfectly suited for invoking the

"wow response" in the audience – the sense of wonder that makes them think "that's so cool!". It is useful to incorporate such moments into a tour, by mentioning things such as:

- The operating temperature of the magnet is -271 degrees Celsius, or 1.9 Kelvin – colder than the ambient temperature of outer space.
- When the LHC ring is cooled from room temperature down to its operating temperature, its length changes by about 80 metres! (This means that care had to be taken in the design to make sure the different parts of the ring remain properly connected to one another throughout the process.)
- On how much energy is contained in mass: If you took just 1 gram of antimatter, and touched that to 1 gram of matter, the resulting explosion would release as much energy as an atomic bomb. (But there's no need to worry – the amount of antimatter produced at CERN is miniscule. Even if we were to annihilate all the antimatter ever made all at once, it would only produce enough energy to power a light bulb for a few seconds.)

11.1.4 The Personal Touch: Humanising CERN Scientists

For a visitor to CERN, their tour guide is often the only scientist they have the chance to meet. I thus like to tell visitors about what life is like at CERN. I explain how many scientists work with the data remotely and how many are physically based on site, and about how one hears five or six languages just walking through the cafeteria.

The visit to the ATLAS control room is a further opportunity to explain some of our roles in more detail. I describe how the machine is manned day and night, over weekends

and holidays, and tell stories about people having to wake up in the middle of the night to rush to the detector because something isn't working. These details help to portray the scientists as ordinary humans, albeit ones who are particularly dedicated to their work.

11.1.5 Weird Anecdotes

Apart from personal stories, anecdotes with a strange twist add a touch of levity and make the tour more memorable. Such stories include the times when the LHC was knocked out by small animals, including the bird who dropped a baguette, and the beech marten who crossed paths with a transformer (its remains are now on display in a museum in Rotterdam).

A stop at the entrance to the AD machine provides an opportunity for a more grisly story – a scene in the movie Angels and Demons, where a scientist's eye is gouged out by a villain in order to gain entry via the retinal scanner. I then point out that this wouldn't work in real life – partly because the retinal quality of a dead eye would be too degraded to pass the scan, and partly because the eye-scanning process requires the eye muscles to focus on a point on the scanner.

11.1.6 Involving the Audience: Asking Questions

Just like in a classroom, a good way to keep the audience engaged is to ask them questions. These should not be overly-difficult questions, but preferably ones that they can figure out the answers to with a little bit of thought. Such questions include:

- **Q.** Why does the LHC shut down for maintanence in the winter (rather than at a different time of year)?

  A. *No, not because we want to go skiing. It's because electricity prices are higher in the wintertime. (The LHC's power consumption is about the same as that of all the households in the canton of Geneva combined.)*

- **Q.** Why did we build the accelerator underground?

  A. *Because it's cheaper to dig a tunnel than to buy up the land above ground.*

- **Q.** Why do we use superconducting electromagnets (as opposed to permanent magnets, or non-superconducting electromagnets)?

  A. *We can't use permanent magnets because they're too weak. And compared to non-superconducting electromagnets, superconducting magnets can generate greater fields because they don't lose energy due to resistance in the coil.*

- At the SC, after visitors have watched a video describing how the accelerator works, I reinforce their understanding by getting volunteers to play the role of protons and electrodes. The electrodes alternate their electric fields, while the protons circle round and round in response, getting faster with each turn.

11.1.7 School Groups

School groups who visit CERN are usually comprised of students who have studied some amount of physics at the high-school level. When conducting these students around the campus, I try to link the things they see to physics concepts that they would have encountered.

For example, on the AD itinerary there are two large electromagnets on display, taken

from the old machine (Figure 11.2). The magnets' coils are visible, and are an excellent opportunity to have the students apply the right-hand rule for Lorentz forces. After explaining the direction that current flows in the coils, I have them figure out the effect that the magnets have on beams of charged particles passing through.

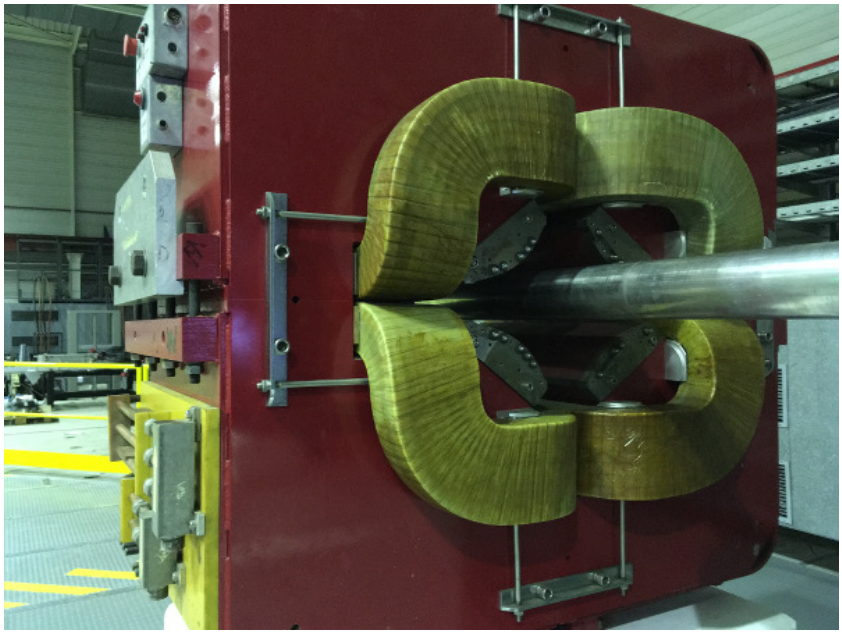


Figure 11.2: Quadrupole magnet on display at the AD.

The experiments at S'Cool Lab are particularly suited for illustrating concepts that students have previously learnt. These links are not always immediately obvious to students. In the electron tube experiment, for example, participants observe that the electron beam in a glass tube gives off an orange colour that doesn't change when the beam intensity or accelerating voltage are changed. By asking them leading questions and giving them hints, I can get them to undergo the "aha" moment where they realise that they knew the reason for this all along – having learnt about atomic energy levels in class.

11.1.8 Broader Applications

Visitors to CERN (especially among the general public) tend to be those who have self-selected for interest in science and physics. While they may be sympathetic to CERN's goal of fundamental research, it doesn't hurt to emphasise some of the technological offshoots of CERN's activities. These include medical imaging devices based on particle detector technology, and research into cancer treatment. Preliminary research at the AD has even shown that antimatter particles have the potential to be used in radiation treatments.

11.2 Coda

Ah, here we are!

```
Finally -- I can't wait to get out and stretch my legs.
```

Hey, we should keep in touch, OK? Let me know how the CERN tours go...

```
Sure thing.  And thanks for telling me about your blog -- I'm excited to
   dig into it.  Oh, and if you ever visit Geneva, let me know and I'll
   show you around CERN!
```

BIBLIOGRAPHY

[1] D. Griffiths, *Introduction to Elementary Particles*. Wiley-VCH, 2008.

[2] I. J. R. Aitchison, "Supersymmetry and the MSSM: An Elementary Introduction", 2005. Lecture Notes. arXiv:hep-ph/0505105.

[3] P. Athron, "Fine Tuning: Standard Model and Beyond". Slides. http://www.physics.gla.ac.uk/~dmiller/doc/FineTuning_forpdfV2.pdf.

[4] S. P. Martin, "A Supersymmetry Primer", 2016. arXiv:hep-ph/9709356.

[5] S. Boutle for the ATLAS, CMS and CDF Collaborations, "Interplay of Top Quark and Higgs Boson Measurements at the Tevatron and LHC", *J. Phys.: Conf. Ser.* **452** (2013) 012007, doi:10.1088/1742-6596/452/1/012007.

[6] CMS Collaboration, "Search for the associated production of the Higgs boson with a top-quark pair", *JHEP* **09** (2014) 087, doi:10.1007/JHEP09(2014)087.

[7] ATLAS Collaboration, "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$TeV with the ATLAS detector", *Eur. Phys. J. C* (2015) 75:349, doi:10.1140/epjc/s10052-015-3543-1.

[8] P. Broks, *Understanding Popular Science*. McGraw-Hill Education, 2006.

[9] C. Wilkinson, "Introduction: A Brief History of Science Communication", January, 2017. University of the West of England, Bristol (Lecture Notes).

[10] C. Wilkinson and E. Weitkamp, *Creative Research Communication: Theory and Practice*. Manchester University Press, 2016.

[11] R. Hunt, *Dictionary of National Biography*. Smith, Elder & Co., 1888.

[12] J. Gillray, "New Discoveries in Pneumatics", 1802. Image. https://en.wikipedia.org/wiki/Humphry_Davy#/media/File:Royal_Institution_-_Humphry_Davy.jpg.

[13] S. Forgan, "A Compendium of Victorian Culture", *Nature* **403.6770** (2000) 596, doi:10.1038/35001134.

[14] B. Rensberger, "Science Journalism: Too Close for Comfort", *Nature* **459** (2009) 1055–1056, doi:10.1038/4591055a.

[15] A. Ridgway, "A Brief History of Science Writing", April, 2017. University of the West of England, Bristol (Lecture Notes).

[16] S. Miller, "Public Understanding of Science at the Crossroads", *Public Understand. Sci.* **10** (2001) 115–120.

[17] B. Wynne, "Misunderstood Misunderstanding: Social Identities and Public Uptake of Science", *Public Understand. Sci.* **1** (1992) 281–304, doi:10.1088/0963-6625/1/3/004.

[18] F. Newport and A. Dugan, "College-Educated Republicans Most Skeptical of Global Warming". Report of Gallup poll. http://news.gallup.com/poll/182159/college-educated-republicans-skeptical-global-warming.aspx?g_source=CATEGORY_CLIMATE_CHANGE&g_medium=topic&g_campaign=tiles.

[19] F. Burnet, "Taking Science to People", 2010. Unpublished guide, University of the West of England, Bristol.

[20] S. Dailler, "Cross section of LHC dipole; Dipole LHC: coupe transversale". Image. https://cds.cern.ch/record/842530.

[21] F. Marcastel, "CERN's Accelerator Complex; La chaîne des accélérateurs du CERN". Image. https://cds.cern.ch/record/1621583.

[22] CMS Collaboration, "CMS Luminosity: Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[23] T. Sakuma and T. McCauley, "Detector and event visualization with SketchUp at the CMS experiment", *J. Phys.: Conf. Ser* **513** (2014) 022032, doi:10.1088/1742-6596/513/2/022032.

[24] A. Dominguez et al., "CMS Technical Design Report for the Pixel Detector Upgrade", Technical Report CERN-LHCC-2012-016, CMS-TDR-11, 2012. doi:10.2172/1151650.

[25] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[26] CMS Collaboration, "CMS tracker performance and readiness for LHC Run II", *Nucl.Instrum.Meth.* **A824** (2016) 67–69, doi:10.1016/j.nima.2015.09.046.

[27] L. Taylor, "About CMS". http://cms.web.cern.ch/content/about-cms.

[28] C. Palmer, "A Search for the Higgs Boson in the H $\to \gamma\gamma$ Channel with CMS", in *DPF-2011 Proceedings*. 2011. arXiv:1109.6805 [hep-ex].

[29] CMS Collaboration, "CMS reconstruction improvement for the muon tracking by the RPC chambers", *JINST* **8** (2013) T03001, doi:10.1088/1748-0221/8/03/T03001.

[30] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003.

[31] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009.

[32] M. Cacciari and G. P. Salam, "The anti-k_t jet clustering algorithm", *JHEP* **0804** (2008) 063, doi:10.1088/1126-6708/2008/04/063.

[33] M. Aldaya et al., "Search for ttH, H→bb decays using the full 2016 data sample", *CMS Draft Analysis Note CMS AN-17-063* (2017).

[34] Y. Coadou, "Boosted Decision Trees", 2016. Slides. https://indico.cern.ch/event/472305/contributions/1982360/attachments/1224979/1792797/ESIPAP_MVA160208-BDT.pdf.

[35] B. P. Roe, H. J. Yang, and J. Zhu, "Boosted decision trees, a powerful event classifier", in *Statistical Problems in Particle Physics, Astrophysics and Cosmology (PHYSTAT 05)*, pp. 139–142. 2006. Proceedings. Edited by Louis Lyons and Muge Karagoz Unel. London, England, Imperial Coll. Press.

[36] N. Chanon, "Statistical Tools in Collider Experiments: Multivariate analysis in high energy physics", 2012. Slides. https://people.phys.ethz.ch/~pheno/Lectures2012_StatisticalTools/slides/Chanon2.pdf.

[37] DMLC, "Introduction to Boosted Trees". http://xgboost.readthedocs.io/en/latest/model.html.

[38] J. Kennedy and R. Eberhart, "Particle Swarm Optimization", *IEEE* (1995) 1942–1948, doi:10.1109/ICNN.1995.488968.

[39] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", *ATL-PHYS-PUB-2011-11, CMS NOTE-2011/005* (2011).

[40] G. Ranucci, "The profile likelihood ratio and the look elsewhere effect in high energy physics", *Nucl.Instrum.Meth.* **A661** (2012) 77–85, doi:10.1016/j.nima.2011.09.047.

[41] S. R. Suhasini, "The Profile Likelihood". Teaching Notes. https://www.stat.tamu.edu/~suhasini/teaching613/profile_likelihood.pdf.

[42] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur.Phys.J.* **C73** (2013) 2501, doi:10.1140/epjc/s10052-011-1554-0.

[43] B. Bray, B. France, and J. K. Gilbert, "Identifying the Essential Elements of Effective Science Communication: What do the experts say?", *International Journal of Science Education, Part B: Communication and Public Engagement* **2:1** (2012) 23–41, doi:10.1080/21548455.2011.611627.

[44] A. Leiserowitz et al., "Global Warming's Six Americas", 2016. Yale Program on Climate Change Communication. http://climatecommunication.yale.edu/about/projects/global-warmings-six-americas/.

[45] C. Roser-Renouf et al., "Engaging Diverse Audiences with Climate Change: Message Strategies for Global Warming's Six Americas", technical report, Yale Project on Climate Change, Center for Climate Change Communication, 2014. doi:10.2139/ssrn.2410650.

[46] P. S. Hart and E. C. Nisbet, "Boomerang Effects in Science Communication: How Motivated Reasoning and Identity Cues Amplify Opinion Polarization About Climate Mitigation Policies", *Communication Research* **39(6)** (2012) 701–723, doi:10.1177/0093650211416646.

[47] L. Fogg Rogers, "Audiences: Audience Engagement Models", January, 2017. University of the West of England, Bristol (Lecture Notes).

[48] D. M. Kahan, H. Jenkins-Smith, and D. Braman, "Cultural cognition of scientific consensus", *Journal of Risk Research* **14(2)** (2011) 147–174, doi:10.1080/13669877.2010.511246.

[49] S. R. Bedrosian et al., "Lessons of Risk Communication and Health Promotion – West Africa and United States", *Centers for Disease Control and Prevention Morbidity and Mortality Weekly Report, Supplement* **65(3)** (2016) 68–74.

[50] Supreme Being, "Particle Fever, or: Hey Everybody! Physics is Awesome!", 2014. Review of Particle Fever. http://www.standbyformindcontrol.com/2014/03/particle-fever-or-hey-everybody-physics-is-awesome/.

[51] E. Weitkamp, "Creating Interesting Science Stories", May, 2017. University of the West of England, Bristol (Lecture Notes).

[52] A. Ridgway, "Writing About the Environment", June, 2017. University of the West of England, Bristol (Lecture Notes).

[53] A. Ridgway, "Writing News Stories", May, 2017. University of the West of England, Bristol (Lecture Notes).

[54] A. Ridgway, "Basic Principles of Science Writing", April, 2017. University of the West of England, Bristol (Lecture Notes).

[55] C. Zimmer, "Carl Zimmer's Brief Guide to Writing Explainers", 2015. https://www.theopennotebook.com/2015/07/07/zimmers-guide-to-explainers/.

[56] T. Radford, "A manifesto for the simple scribe – my 25 commandments for journalists", 2011. https://www.theguardian.com/science/blog/2011/jan/19/manifesto-simple-scribe-commandments-journalists.

[57] J. E. Sundermann and T. Göpfert, "KinFitter – A Kinematic Fit with Constraints". Code Documentation. http://iktp.tu-dresden.de/~goepfert/KinFitter.pdf.

[58] P. Avery, "Vertexing and Kinematic Fitting, Part I: Basic Theory", 1998. Lecture Notes. http://www.phys.ufl.edu/~avery/fitting/kinfit_talk1.pdf.

[59] S. Yaschenko, "Kinematic Fitting – A powerful tool of event selection and reconstruction", 2011. Slides. http://www-hermes.desy.de/notes/pub/TALK/yaschenk.ColloqGlasgow.pdf.

[60] L. Winstrom, G. N. Kaufman, and J. Thom, "Object Reconstruction in Collider Events Containing Top Quarks", 2014.

[61] B. A. Betchart, R. Demina, and A. Harel, "Analytic solutions for neutrino momenta in decay of top quarks", *Nucl. Instrum. Meth.* **A736** (2014) 169–178, doi:10.1016/j.nima.2013.10.039.

[62] F. James and M. Winkler, "Minuit User's Guide", 2004.

[63] F. James, "Minuit Tutorial: Function Minimization", 2004. Reprinted from the Proceedings of the 1972 CERN Computing and Data Processing School, Pertisau, Austria, 10-24 September, 1972.

[64] A. Ridgway, "Introduction to Writing for Different Audiences", May, 2017. University of the West of England, Bristol (Lecture Notes).

[65] A. Ridgway, "Principles of Writing Online", May, 2017. University of the West of England, Bristol (Lecture Notes).

[66] S. M. Wood, "Harmonic Series". http://www.musiccrashcourses.com/lessons/harmonic_series.html.

[67] Music Technology Group, Universitat Pompeu Fabra, "sms-tools". Sound analysis/synthesis tools for music applications. https://github.com/MTG/sms-tools.

[68] R. Muehleisen, "Simple Traveling Plane Wave", 2006. http://mypages.iit.edu/~muehleisen/acs_demos/wave_animations/planewave.html.

[69] M. Erdmann, B. Fischer, and M. Rieger, "Jet-Parton Assignment in $t\bar{t}H$ Events using Deep Learning", *JINST* **12** (2017) P08020, doi:10.1088/1748-0221/12/08/P08020.